Subcategorization Acquisition and Classes of Predication in Urdu

Dissertation zur Erlangung des akademischen Grades eines Doktors der Philosophie

vorgelegt von

Ghulam Raza

an der Universität Konstanz

Geisteswissenschaftliche Sektion
Fachbereich Sprachwissenschaft

Tag der mündlichen Prüfung: 30 November 2011

Referentin (Chair): Dr. Heike Zinsmeister

Referentin: Prof. Dr. Miriam Butt

Referent: Prof. Dr. Rajesh Bhatt

Dedicated to *Abbaa Saeen* and *Ammaan Huzoor*

Acknowledgements

Hinrich Schütze was my first mentor. He supervised my PhD research for more than two years. He spared enough time to let me discuss different topics of computational linguistics with him. It was a wonderful experience to learn the concepts of information retrieval from a very good book which Hinrich Schütze is a co-author of. In the beginning, he assigned me the task of improving the ranking of German parses by encoding lexical features. Although the results were not positive, it was a great opportunity for me to learn. Since I did not have a good knowledge of the German language, I shunned experimenting on German corpora further and changed my research topic to the automatic acquisition of subcategorization information of verbs from an Urdu corpus under the supervision of Miriam Butt.

She exposed me to the Lexical Functional Grammar framework in which I later analyzed some Urdu constructions. Exploring interesting syntactic phenomena in Urdu language and sharing them with her has been a source of joy and happiness for me. She spent a lot of her precious time on commenting and correcting many drafts of my presentations, papers and the thesis. My knowledge of English and linguistics should now definitely be better than it was three years before. Tikaram Poudel is the person from whom I learnt how linguists think and react to empirical observations in a language. His style of discussing a language and linguistics was inspiring. Unfortunately, he left Konstanz some months after I reached here.

Acknowledgements are due to Helmut Schmidt, Martin Forst, Aoifa Cahill and Sabine Schulte im Walde who always readily responded whenever I asked them for some help during my stay at the University of Stuttgart. Kati Schweitzer was so nice to happily make a copy of a statistics book for me at times when I did not know how to use a copy machine. Alice Davison took the headache of sending me the printed copy of one of her articles from the USA when I requested her to send just an electronic copy.

The colleagues who have been helpful at Konstanz are: Jaouad Mousser, Melanie Seiss, Nette Hautli, Qaiser Abbas, Sebastian Sulger, Tafseer Ahmed

and Tina Bögel. In particular, Jaouad Mousser provided me a Latex template to write the thesis; Melanie Seiss proved an angel on many occasions when I was stuck with Latex problems; Nette Hautli translated the summary of the thesis in German. Rajesh Bhatt commented on the first draft of the thesis and identified some errors.

The higher education commission (HEC) of Pakistan was generous to award me the PhD scholarship and the Pakistan Institute of Engineering and Applied Sciences (PIEAS), where I am employed in Pakistan, was kind to grant me the study leave.

Thanks to all!

Transcription Scheme

Consonants//	Orthographic	Phonetic (IPA)	Transcription
	ب	b	b
	پ	p	p
	ت	t̪	t
	ٹ	ʈ	ṭ
	ث	θ	s
	ج	dʒ	j
	چ	tʃ	c
	ح	ħ	h
	خ	x	x
	د	t̪	d
	ڈ	ɖ	ḍ
	ذ	ð	z
	ر	r	r
	ڑ	ɽ	ṛ
	ز	z	z
	ژ	ʒ	y
	س	s	s
	ش	ʃ	ʃ
	ص	sˤ	s
	ض	dˤ	z
	ط	tˤ	t
	ظ	ðˤ	z
	ع	ʕ	ɑ
	غ	ɣ	ɣ
	ف	f	f
	ق	q	q
	ک	k	k
	گ	g	g
	ل	l	l
	م	m	m
	ن	n	n
	و	ʋ	v
	ہ	h	h
	ی	j	y
	ھ- (Aspiration)	-ʰ	-ʰ

Vowels//	Orthographic	Phonetic (IPA)	Transcription
	اَ	ə	ɑ
	اِ	ɪ	ı
	اُ	ʊ	ʊ
	ا	a:	a
	اے	e	e
	اَے	æ	ɛ
	اِی	i:	i
	او	o	o
	اَو	ɔ	ɔ
	اُو	u:	u
	ں (Nasalization)	˜	˜

Morphemic Glosses

Gloss	**Meaning**
1	first person
2	second person
3	third person
M	male
F	female
Neut	neutral
Sg	singular
Pl	plural
Pres	present tense
Past	past tense
Fut	future
Nom	nominative
Acc	accusative
Dat	dative
Erg	ergative
Gen	genitive
GenR	reflexive genitive
Refl	reflexive
Pron	pronoun
RelP	relative pronoun
Abl	ablative
Inst	instrumental
Loc	locative
Temp	temporal
Obl	oblique form
Dir	direct form
Inf	infinitive
Perf	perfect aspect
Imperf	imperfect aspect
Subjn	subjunctive
Prog	progressive
Caus	causative
Ez	*ezafe*
Emp	emphatic
Conj	conjunction

Abbreviations

NLP	Natural Language Processing
SCF	Subcategorization Frame
CLC	Case Clitic and Complementizer Combination
SASU	Subcategorization Acquisition System for Urdu
XLE	Xerox Linguistic Environment

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Syntactic structure in most recent syntactic theories is projected from the lexicon which provides access to subcategorization information for different predicators. Recording correct and comprehensive subcategorization information in the lexicon is essential for the development of a robust and accurate parser. In addition to parsing, subcategorization information is also useful in other natural language processing tasks. This thesis centers around the identification and exploration of different types of predicators and their subacategorization patterns in a South Asian language, Urdu. A broad-coverage robust grammar of Urdu language in Lexical Functional Grammar (LFG) framework is being developed in Urdu ParGram project (see section 1.5) at the University of Konstanz and efforts are being made to develop lexical resources for the Urdu language. The research presented in this thesis will ultimately help build a computational lexicon for the Urdu language.

In this introductory chapter, notions of predication and subcategorization are briefly described and it is argued why it is necessary to explore a language individually for subcategorization information.

1.1 Predication

Predication is saying or attributing something of something else. Sentences in a language are instances of predication. In traditional grammar, a sentence is assumed to be composed of a *subject* and a *predicate*. The subject is the entity about which some expression is uttered and the predicate, from a logical perspective, is an expression that can be true of something or from the perspective of traditional grammar, all that is said of the subject.

In contemporary linguistics, a sentence or a predication is semantically analyzed into a predicator and its arguments. Arguments can be entities (referential expres-

sions) or ideas which can be talked about for some property or some relation. In a simple declarative sentence, predicator is the word or group of words which does not belong to any of the referring expressions, however, it renders some property or some relation/process in which the referents of referential expressions are involved (Hurford & Heasley, 1983).

(1) a. Ali came.
Argument Predicator

b. Ali pinched an elephant.
Argument Predicator Argument

c. Ali gave a book to Nida.
Argument Predicator Argument Argument

Some instances of verbal predication in English are given in (1). Degree or valency of the predicator is a number of arguments it has to have for a meaningful statement/expression. The verb *come* in (1a) is a one-place predicator as it takes a single argument; the *pinched* in (1b) and the verb *gave* in (1c) are a two-place predicator and a three-place predicator, respectively, because they take two and three arguments, respectively. Not only noun phrases (or pronouns) fill the positions of arguments but other constituents or even clauses can also act as arguments of the predicator in a sentence across languages. Two argument positions of the predicator *know* in (2a) are filled by two nouns and in (2b) one position is filled by a noun and other by a clause.

(2) a. Nida knows Ali.
Argument Predicator Argument

b. Nida knows that Ali came.
Argument Predicator Argument (clause)

Since a predicator shows the relation between arguments, and also between arguments and the real world, not only verbs but other parts of speech can also function as predicators. Nouns, adjectives and prepositions can also function as predicators in addition to (main) verbs, but conjunctions and articles cannot. Some examples of non-verbal predication in English are given in (3).

(3) a. Nida is beautiful.
Argument Predicator

b. Ali is a genius.
Argument Predicator

c. Qum is in Iran.
Argument Predicator Argument

The adjective *beautiful* in (3a) is a predicator which relates *Nida* to the external world. This instance of predication predicates beauty (a property) of Nida. The noun *genius* in (3b) and the preposition *in* in (3c) also function as predicators. Prepositions like *in* are two-place predicators. The verb *is* in these examples does not function semantically, but is used grammtically to link the predicator with its first argument. In many other languages like Russian and Arabic (4) no such word is required, rather the predicator and its argument are simply juxtaposed.

(4) a. Moskva gorod.
Moscow city
'Moscow is a city.' (Russian: Raptschinsky 1946:11)

b. الرجل كريم
ɑl-rɑjʊlʊ kɑrɪmʊn.
the-man.Nom generous.Nom
The man is generous.' (Arabic: Attia 2008)

Although the majority of nouns are one-place predicators, some also function as two-place predicators (5a). Inherently relational nouns (Hurford & Heasley, 1983) like *brother, neighbor*, act as two-place predicators in sentences. Some adjectives are also two place predicators, as in (5b).

(5) a. Ali is a genius at story-writing.
Argument Predicator Argument

b. Nida is certain of her success.
Argument Predicator Argument

It is worth mentioning here that the term *predicate* in contemporary linguistics is meant to be a word or a group of words in a language that potentially can act as a predicator. Predicator identifies the semantic role played by a particular word or words in a particular sentence whereas predicate identifies elements in a language

independently of particular example sentences. In the sentence, *A tall, pretty woman entered the hotel*, only the word *entered* is the predicator and hence predicate too. However, the other words *tall, pretty, woman and hotel* are also predicates as they can function as predicators in other sentences (*She is tall, She is pretty, She is a woman, That building is a hotel*). From here onwards, I will use the terms predicator and predicate alternatively.

1.2 Subcategorization

Subcategorisation is concerned with the specification of arguments which are salient to the local context of a predicate i.e. they are semantically selected by the predicate. Syntactically diverse behavior of verbs is explained in terms of subcategorizaion. Consider, for example, the two verbs *eat* and *put* in (6) and (7).

(6) a. Ali ate an apple.

b. Ali ate.

(7) a. Nida put the pen in the drawer.

b. * Nida put in the drawer.

c. * Nida put the pen.

d. * Nida put.

The verb *eat* in English is used either transitively taking an NP complement (6a) or it is used intransitively (6b). However, the verb *put* takes an NP-PP complement (7a) and it does not allow only a PP complement (7b) or only an NP complement (7c). It cannot be used intransitively (7d) in contrast with *eat*, rather it obligatorily requires three arguments: a subject, an object and an oblique argument.

Subcategorization information is characterized in terms of syntactic frames called subcategorization frames (SCF) that specify the number and type of arguments that a particular predicate requires. So, the set of arguments a predicate takes is called its subcategorization frame. It is possible, as we observed in (6), that a single predicate can take a variable set of arguments in different situations. In that case the predicate is said to have more than one subcategorization frame. Knowing about all possible frames of a predicate is very important for various natural language processing tasks. For example, Briscoe & Carrol (1993) parsed unseen test data on a parsing system utilizing

a lexicalist grammatical framework and noted that half of the parse failures were due to inaccurate subcategorization information. To project an accurate syntactic structure of any language, most of the grammar formalisms today require comprehensive lexicons having accurate information about the predicate subcategorization. Subcategorization information can also be integrated in dictionaries (Evans & Kilgarrriff 1995; Gahl 1998). In psycholinguistic research this information is used for approximating lexical preferences (Lapata & Keller 1998; Lapata *et al.* 2001) and it can also provide the empirical basis for linguistic theory (Levin, 1993).

The participants involved in the event described by verbs constitute the the subcategorization frames of verbs. Every participant has some semantic or thematic role in the event (Dowty, 1991; Fillmore, 1968; Gruber, 1976). The most frequently used thematic roles and their definitions are listed below.

- **Agent** The participant that performs or causes some event. It is often with volition.
- **Patient** The affected participant.
- **Theme** The participant whose position or state is changed.
- **Experiencer** The participant that experiences or becomes informed in the event.
- **Source/Goal/Location** The participant which entails the starting or ending positions of the event or merely the location.
- **Recipient/Beneficiary/Maleficiary** The participant that receives (recipient) or benefits/suffers form (beneficiary/maleficiary) from something in the event.
- **Instrument** The participant used as a means for doing something.

For example, the verb *come* would have only one theme argument in its subcategorization frame, the verb *cut* would have two arguments: agent and theme and the verb *give* would have three arguments: agent, theme and goal. However, every argument in the syntactic frame of a verb does not necessarily have some semantic role. Consider the example sentences in (8), adapted from Korhonen (2002).

(8) a. **John** seems to drive a Ferrari.

b. **It** seems to annoy Tim that John drives a Ferrari.

c. **John** tries to drive a Ferrari.

d. * **It** tries to annoy Tim that John drives a Ferrari.

The subject of the raising verb *seem* in (8a) is contentful and in (8b) it is the pleonastic. Syntactically the subject argument of raising verbs is not assigned a semantic role. The subject *John* in (8a) is actually the raised subject of the verb *drive* in the clausal argument of the main verb *seem* (Bresnan, 1982). The verb *seem* is considered to be a subject raising verb and takes only a clausal argument, hence is a one-place predicate semantically. There are some control verbs in English which do not take non-thematic arguments and are generally distinguished as equi verbs. All subcategorized dependents of equi verbs are assigned semantic roles on the other hand. The verb *try* is an example of an equi verb. Its subject is always contentful. The subject *John* in (8c) is the subject of both the verb *try* and and the other verb *drive* of the clausal argument of the verb *try*. This is the reason that the equi verbs cannot have pleonastic subject (8d). The verb *try* is a two-place predicate semantically and syntactically. Mapping semantic arguments of verbs to syntactic arguments in theory is called 'linking' (Bresnan, 1982).

Subctegorization information of predicates collected for one language, say English, cannot be generalized for all languages as languages may differ in many respects. For one, semantically similar versions of verbs are not found in other languages. For example, the inchoative version of the transitive verb *cut* is not found in English but is found in Urdu. This is illustrated in (9)–(10).

(9) a. Ali cut the apple.

b. * The apple cut.

(10) a. علی نے سیب کاٹا.

ɑli=ne	seb	kaṭ-a
Ali.M.3Sg=Erg	apple.M.3Sg	cut-Perf.M.3Sg

'Ali cut the apple.'

b. سیب کٹا.

seb	kɑṭ	gɑ-ya
Apple.M.3Sg	cut-Perf.M.3Sg	

'The apple got cut.'

Secondly, an argument of a verb in a language can be implicitly included in the semantics of the verb in another language. For example, in the verb *stab* 'pierce with a knife' of English the semantics of the object *knife* is implicitly included in the verb. There is no equivalent verb in Urdu. Instead the verb *mar* 'hit' is used with the object argument *knife* (11).

(11) a. The robber stabbed Ali in the head.

b. ڈاکو نے علی کو سر میں چاقو مارا.

ḍaku=ne	ɑli=ko	sɑr=mẽ	caqu	mar-a
robber.M.3Sg=Erg	Ali.M.3Sg=Dat	head.M.3Sg=Loc.in	knife	hit-Perf.M.3Sg

'The robber hit Ali in the head with a knife.

Thirdly, verbs of similar semantics can take different types of arguments in different languages. Consider the verb *meet* and its corresponding verb *mɪl* in Urdu in (12). Although they denote the same semantics, in English an accusative object is needed whereas in Urdu a comitative object is required.

(12) a. Ali met Nida.

b. علی ندا سے ملا.

ɑli	nɪda=se	mɪl-a
Ali.M.3Sg	Nida.F.3Sg=Com	meet-Perf.M.3Sg

'Ali met Nida.

Fourthly, there could be the case that a verb is used in an additional sense in one language but not in another language. The arguments in the additional sense might not be the same. For example, consider again the verb *mɪl* in Urdu, which is also used in the sense of *get* and in that case a dative subject is required (13).

(13) a. * To Ali met the letter.

b. علی کو خط ملا.
ɑli=ko xɑt mɪl-a
Ali.M.3Sg=Dat letter.M.3Sg meet-Perf.M.3Sg
'Ali got the letter.

Fifthly, argument alternations found in one language for a given verb might not be found in another language. For example, for the verb *give* in English the second complement could be either in the form of a prepositional phrase or a dative. In Urdu, however, the indirect object of *de* 'give' is always dative (14).

(14) a. Ali gave a pen to Nida/Ali gave her a pen.

b. علی نے اس کو قلم دیا.
ɑli=ne ʊs=ko qɑlɑmt dɪ-ya
Ali.M.3Sg=Erg 3Sg=Dat pen.M.3Sg give-Perf.M.3Sg
'Ali gave her a pen.

Due to these facts, the subcategorization lexicon of one language cannot be directly built by using the lexicon of some other language. Instead, each language must be investigated individually and a comprehensive subcategorization information for its lexicon must be acquired on a language by language basis.

1.3 Subcategorization in LFG

Lexical-Functional Grammar (LFG) is a lexicalist, constraint-based, non-transformational theory of grammar. Different kinds of linguistic information are modeled in different representations. Different representations are formally related by correspondence functions. There are mainly two kinds of representations (Austin, 2001; Dalrymple, 2001; Kaplan & Bresnan, 1995): (i) constituent-structure (c-structure) encodes the linear order and constituency in form of trees and (ii) functional-stucture (f-atructure) represents the grammatical functions of a proposition in form of an attribute-value matrix

(avm). F-structure represents a sentence in terms of subject, object as primitives. Languages may differ considerably at the c-structure level but the f-structure representations are generally very similar across languages (Butt *et al.*, 1999).

Linguistic theories differ with respect to the exact type of subcategorization information provided in the lexical entry of a predicate. For example, the control links of raising and equi verbs which are handled in terms of syntactic principles or rules in Government and Binding (GB) theory can be lexically specified in LFG.

The grammatical functions used in LFG are: SUBJect, OBJect, OBJect$_\theta$, OBLique$_\theta$, COMP, XCOPM, ADJUNCT, XADJUNCT. The core function are SUBJ and OBJ. OBJ$_\theta$ and OBL$_\theta$ are thematically restricted functions. OBJ$_\theta$ corresponds to the secondary object with some specialized thematic role and OBL$_\theta$ to adpositional phrase with some specialized thematic role. All functions are governable by the verb except ADJUNCT and XADJUNCT, which are adjuncts.

The clausal expressions which are controlled from within are assigned closed functions, COMP or ADJUNCT depending upon whether they are complement or adjunct of the main clause. Likewise the clausal expressions which are controlled from outside/externally are assigned open functions XCOMP or XADJUNCT. Examples of closed and open grammatical functions are give in (15)–(16), taken from (EAGLES, 1996).

(15) a. John believes [*that Bill is a genius*]$_{\text{COMP}}$.

b. [*John being angry*]$_{\text{ADJUNCT}}$, Mary left.

(16) a. John wants [*to be a genius*]$_{\text{XCOMP}}$.

b. [*Being angry at John*]$_{\text{XADJUNCT}}$, Mary left.

Grammatical functions are placed in correspondence with thematic roles to obtain the lexical form with syntactically subcategorized functions. As an example, the predicate-argument structure for the verb *break* and its lexical form is given in (17).

(17) a. predicate argument structure: ‘break<agent, theme>’

b. grammatical function assignment: ((SUBJ),(OBJ))

c. lexical form:

	‘break<	agent	theme	>’
c. lexical form:		|	|	
		SUBJect	OBJect	

Grammatical functions in LFG are assigned as per traditional tenets of the grammatical theory. Every intransitive verb has a SUBJ function and a transitive verb has both SUBJ and OBJ functions. Syntactic subcategorization information is specified in the lexicon which has to appear in the f-structure. The feature ‘PRED’ is used for the semantic form. The lexical entires of some verbs of English are shown in the following:

- come: PRED ‘come $< SUBJ >$’
- cut: PRED ‘cut $< SUBJ, OBJ >$’
- eat: PRED ‘eat $< SUBJ, (OBJ) >$’
- give: PRED ‘give $< SUBJ, OBJ, OBJ_\theta >$’
- seem: PRED ‘seem $< XCOMP > SUBJ$’

The optional OBJ argument of the verb *eat* is shown in parenthesis and the SUBJ argument of the raising verb *seem* is written out of angular brackets because it is non-thematic. The f-structure representation of ‘Ali came yesterday’ is shown in Figure 1.1. At f-structure, values of attributes can be sets for unbounded number of elements, for example, adjuncts and elements in coordination. Due to this reason the ADJUNCT in Figure 1.1 is placed in curly brackets (Dalrymple, 2001).

$$\begin{bmatrix} \text{PRED} & \text{‘come}\left\langle \text{SUBJ} \right\rangle\text{’} \\ \text{TENSE} & \text{Past} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{‘Ali’} \\ \text{NUM} & \text{sg} \end{bmatrix} \\ \text{ADJUNCT} & \left\{ \begin{bmatrix} \text{PRED} & \text{‘yesterday’} \end{bmatrix} \right\} \end{bmatrix}$$

Figure 1.1: An f-structure representation

For the proper noun ‘Ali’ in the f-structure PRED and NUM are attributes and ‘Ali’ and ‘sg’ are corresponding values. LFG does not restrict a fixed number of features or values to be included in the f-structure, rather it is upto the choice of a grammar writer (Butt *et al.*, 1999). Attributes can share values. This feature helps handle the phenomenon of raising in LFG, as shown in Figure 1.2 where the SUBJ of main verb and that of XCOMP share a single value.

An f-structure in LFG is constrained by the following three conditions (Bresnan, 1982), the first two of which refer to subcategorization frames.

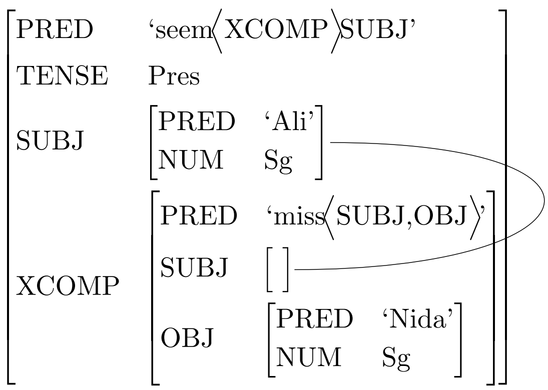


Figure 1.2: Illustration of XCOMP function

- Completeness
- Coherence
- Uniqueness

The **completeness** condition entails that all governable functions of a predicate mentioned in the lexical entry should be present in the f-structure.

(18) a. lift: PRED ‘lift $< SUBJ, OBJ >$’

b. Ali lifted the chair.

c. * Ali lifted.

The lexical entry for the verb *lift* is shown in (18a). This verb subcategorizes for SUBJ and OBJ arguments. The sentence in (18b) has both arguments, so its f-structure would be well formed. However, the sentence in (18c) would be ruled out by the grammar as it lacks the information specified in the lexical entry.

The **coherence** condition ensures that no argument should be part of f-structure which is not governed by the predicate i.e. is not mentioned in the lexical entry of the verb.

(19) a. arrive: PRED ‘arrive$< SUBJ >$’

b. Ali arrived.

c. * Ali arrived Nida.

The verb *arrive* takes only SUBJ argument as is mentioned in the lexical entry (19a). The invalid sentence in (19c) would not be parsed as it has an argument which is not governed by the verb.

Finally, the **uniqueness** condition states that every attribute must have a single value. It makes it possible to meet agreement conditions for different features in the f-strucure representation.

So, the LFG framework has formal notations for encoding subcategorization information of predicates in lexicon and while parsing, subject to different conditions on grammar rules, the sentences with incorrect subcateogrization are ruled out.

1.4 ParGram

The Parallel Grammar (ParGram) project began in 1994, initially to develop grammars of three European languages: English, German and French using the LFG formalism (Butt *et al.*, 1999). Later on three other languages: Japanese, Norwegian and Urdu were also added (Butt *et al.*, 2002). At present grammars for Arabic, Chinese, Bahasa Indonesian, Malagasy, Turkish, Welsh and Murrinh-patha are also being developed in the LFG framework. The ParGram project uses the XLE parser and grammar development platform (Maxwell & Kaplan, 1993). The parallelism across languages is encoded at the f-structure level of abstraction within LFG. Language particular differences in linear word order, surface morphological vs. syntactic structures, and constituency are encoded at the c-structure.

As an example for the parallel analysis of language constructions (Butt *et al.*, 2002) consider the interrogative sentences form German and Urdu in (20)–(21).

(20) Was hat John Maria gegeben?
What has John Mary given.Perf?
'What did John give to Mary?' (German)

(21) a. جون نے میری کو کیا دیا؟
jɔn=ne mɛri=ko kɪya dɪ-ya?
John=Erg Mary=Dat what give-Perf.M.3Sg?
'What did John give to Mary?'.

b. jɔn=ne kɪya mɛri=ko dɪ-ya?

c. jɔn=ne mɛri=ko dɪ-ya kɪya?

In German, like in English, the interrogative word comes in the first position (20) with the finite verb in the second position. In Urdu, however, the interrogative word can appear in a number of positions in an interrogative sentence as is shown in (21a–c). Due to differences in word order the c-structures of the sentences will be different, however the f-structure representation for interrogatives of both languages would be identical as is shown in Figure 1.3.

$$\begin{bmatrix} \text{PRED} & \text{'give}\left\langle\text{SUBJ, OBJ, OBL}\right\rangle\text{'} \\ \text{FOCUS-INT} & \begin{bmatrix} \text{PRED} & \text{'pro'} \\ \text{PRON-TYPE} & \text{'int'} \end{bmatrix}_1 \\ \text{TENSE} & \text{'past'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'John'} \\ \text{NUM} & \text{'Sg'} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \ \end{bmatrix}_1 \\ \text{OBL} & \begin{bmatrix} \text{PRED} & \text{'Mary'} \\ \text{NUM} & \text{'Sg'} \end{bmatrix} \\ \text{STMT-TYPE} & \text{int} \end{bmatrix}$$

Figure 1.3: f-structure representation

The grammars for different languages are developed on different sites and are consolidated at project meetings held twice a year. In the meetings, analyses of sample sentences are compared and the justification for any differences is discussed. If there is more than one way to implement some analysis, then conceptual simplicity and computational efficiency are taken into account in choosing an analysis.

1.5 Subcategorization lexicons

Many lexicons have been developed either manually, automatically or semi-automatically for different languages. These lexicons differ with respect to the number of entries and

the extent and the structure of information. Approaches to develop verb subcategorization lexicons are distinguished with respect to the following aspects:

- Theoretical basis of the lexicon
- Meeting some specific application
- Number of levels of linguistic description
- The model of the lexical entries
- Augmented information for subcategorization pattern, word sense and part of speech

Some lexicons are developed for a specific application and, sometimes, some linguistic theory is taken as a basis for the lexicon development. The number of arguments, syntactic category and functional role of arguments, lexical selection and morphosyntactic constraints are important aspects among others to be considered while developing a lexicon. For English, many subcategorization lexicons have been developed manually. Among them the largest one is VerbNet (Kipper-Schuler, 2005) which has been constructed on the basis of Levin (1993)'s verb classification. Many efforts have also been made in terms of the automized building of such lexicons for English. These efforts comprise the acquisition from a raw English corpus (Brent, 1991, 1993) and from an annotated part of speech (POS) tagged corpus (Manning, 1993; Ushioda *et al.*, 1993) and from a corpus parsed partially or fully (Briscoe & Carrol, 1997; Kinyon & Prolo, 2002; O'Donovan *et al.*, 2005). Other major lexicons are listed in the following (EAGLES, 1996):

ACQUILEX: This multilingual (Dutch, English, Italian and Spanish) lexicon was developed within the Acquilex project (Sanfilippo, 1993). It was heavily based on Categorial Grammar model and a sign-based approach was used for lexical representation. Unification categorial grammar (UCG) is a radically lexical theory of grammar, so the maximum lexical information was put in the lexical entires. For other frameworks the information could be processed automatically. The predicate arity is logically specified as a conjunction of formulae whose main predicates are thematic relations.

COMLEX: This American English lexicon was built by the University of Pennsylvania for the Linguistic Data Consortium. A typed feature structure formalism was used (Rohen Wolff *et al.*, 1994). A list of detailed definitions of subcategorization features are given in the dictionary. The verb frames are mentioned in terms of constituents such as NP, PP, ADJP, etc as well as in terms of grammatical relations; the

corresponding ones are linked with a common index followed by them. The number of arguments can be induced from the number of complements specied at the level of grammmatical structure.

EUROTA: This was developed for generation and analysis in the Eurotra Machine Translation system (ten Hacken *et al.*, 1991). Four different levels of linguistic description are implied in this system: the morphological level, the configurational or constituent level, the relational syntactic level and the interface or semantic level. Each level has its own grammar and its own lexicon. The last three levels correspond to the issue of verbal subcategorisation. At both levels of relational and interface structure the number of obligatory arguments/complements are specified.

GENELEX: This full-scale application- and theory-independent lexicon was developed within the Genelex project (GENELEX, 1993). The architecture of the lexicon is conceptually based on the entity-relationship model. The theory independent lexical information can be derived from the lexicon. The positions and syntactic realizations of complements are specified with all sorts of restrictions. Complementation patterns can be tailored by lexicographers to meet specicific needs. There is some position in lexical entry to mention the number of arguments.

ILCLEX: This is an Italian lexicon. It was integrated with a robust wide-coverage corpus grammar, which collects information extracted from machine readable sources such as dictionaries and corpora (Vanocchi *et al.*, 1994). Lexical representations are theory-independent with three information levels: pattern rules, patterns and subcategorization frames. The grammatical contexts in which the verb can occur are mentioned at a third level. The number of arguments in some frame are specified by a feature which can take a numerical value.

LDOCE: The Longman Dictionary of Contemporary English (Procter, 1987), has been used for the construction of several computational lexicons (see, for example, Boguraev & Briscoe 1989. This dictionary describes subcategorization information in terms of types and number of complements in addition to part of speech information. Alternative subcategorizations are mentioned within the same entry.

PLNLP: This Italian lexicon was used by the PLNLP grammar originally developed for a style-checking application and was later used for different other computational tasks. It is a broad coverage lexicon consisting of a simple list of words and their featural information. In LODOCE and PLNLP, the subcategorization information is registered in terms of conventional grammatical codes, that is, T and I for transitive and intransitive verbs and the number of arguments that are inferred from such codes.

IMSLex: This is a German lexicon which covers valency information in addition to inflection and word formation of base forms. From this lexicon special lexical data

can be derived to be used in information retrieval and information extraction and other applications.

No complete subcategorization lexicon for Urdu has been developed manually or automatically upto now. Many dictionaries of Urdu exist and recently, the Urdu Lughat Board in Pakistan has published twenty-one volumes of a large Urdu dictionary. However, all these dictionaries lack subcategorization information. In the Center for Research in Urdu Language Processing (CRULP) in Pakistan some efforts have already been made towards Urdu lexicon development (Ijaz & Hussain, 2007) without much focus on subcategorization information. It is very hard to develop a lexicon with subcategorization information manually in a reasonable amount of time. In this thesis, a method is therefore proposed to extract subcategrization information of verbs from a raw Urdu corpus.

1.6 Contribution of the thesis

The objective of this thesis is to identify and explore subcategorization patterns of different classes of predicators in Urdu. The thesis covers the following points:

- Are the existing subcategorization acquisition strategies for other languages applicable to acquire subcategorization frames of verbs in Urdu? If no, what challenges and problems are there?
- Identification of what types of arguments a predicator can take in Urdu
- Given the explored challenges, could a scheme be developed to extract subcategorization information of verbs automatically from a raw Urdu corpus?
- The devising of an algorithm; implementing and evaluating it and reporting the results

During this core research some interesting phenomena regarding predication and subcategorization have been observed and have been further explored. The subcategorization information which the developed system is unable to extract is empirically investigated. In this regard the following points are also crucially included in the thesis.

- Semantic realizations and syntactic behavior of the verb *ho* ‘be’ in Urdu
- The type of arguments deverbal elements can take in Urdu and their positions in the c-structure relative to their heads. The c-structure is discussed in the thesis after having found an unexpected order of arguments relative to their heads in noun phrases.

- Adpositions in Urdu play an important role in Urdu predication. This thesis analyses also different adpositions in Urdu, their origin, the arguments they take and their c-structure.

1.7 Outline of the thesis

The thesis aims at exploring the arguments of different types of predicators, that is, verbs, adjectives, nouns and adpositions in Urdu. Different types of lexical and syntactical predicators are also identified. A method is proposed to acquire aubcategorization frames of verbs in Urdu from a raw Urdu corpus. A raw Urdu corpus is used for this purpose as we do not have refined resources e.g., a part of speech tagged corpus or an Urdu tree bank. In Chapter 2, different challenges to the subcategorization acquisition from a raw Urdu corpus are presented. Established structural cues for subcategorization acquistion cannot work for Urdu as it is fairly free word order language. The arguments of verbs can appear in different orders before the verb in a sentence. Participants of a verbs are usually marked for case in Urdu. I have therefore worked out a system of acquisition in which lexical cues of case clitics are used to find specific case phrases for a target verb and the syntactic frames of the verb are inferred indirectly form the case clitics combinations for the verb. The scheme developed for extracting frames of verbs will be presented in Chapter 3.

The verb *ho* 'be' in Urdu is explored for its subcategorization frames in Chapter 4. This verb shows diversified syntactic and semantic behavior. It is used both as an intransitive verb and a copular verb and can act both as a stative and as a dynamic verb. Syntactic tests for distinguishing its stative and dynamic uses are developed. The copular *ho* is also used to encode the notion of possession in Urdu. In addition to this, the same verb is used in the construction of participial adjectives. In different constructions in which the verb *ho* is used as a linking verb, the subject and predicate distinction is not straightforward. It is argued that the position of the participant matters for considering it as the subject or the predicate.

Arguments of deverbal adjectives and nouns are explored in Chapter 5. When the arguments of deverbal adjectives and deverbal nouns come together in noun phrases they can scramble within the NP before their heads. Due to this scrambling discontinuous constituents are generated in NPs. This phenomenon is evidenced by putting examples from news corpora. Its syntactic explanation and an implementation in LFG framework is presented. .

Verbs can take prepositional phrases as their arguments or adjuncts which are discussed in detail in Chapter 6. From the corpus-based exploration I present a variety

of adpositions in Urdu. A model of spatial adpositions in general is posited in LFG by introducing lexical-semantic features. Classes of adpositions in Urdu based on their syntactic complements are explored and complex adpositions are analysed and modeled in the LFG framework.

The first half of the thesis (Chapters 1–3) thus presents a system for automatically acquiring subcategorization information of verbs in Urdu from a raw Urdu corpus. It was hard to extract the subcategorization information of the verb *ho* 'be' through the developed system because this verb exhibits diversified syntacto-semantic behavior. The subcategorization frames of this verb and other predicators like nouns and adpositions are therefore investigated empirically in the second part of the thesis (Chapters 4–6). Chapter 7 concludes the thesis.

Chapter 2

Urdu verbs and challenges for lexical acquisition

Urdu has a variety of verbal patterns. Major constituents entailing participants of a verb in a clause can be scrambled freely among each other. The verb itself usually comes last. The core arguments or complements of a verb are generally overtly marked for case (Butt & King, 2005). Adjuncts are either case marked phrases or adpositional phrases. In this chapter different types of verbs in Urdu are described and the challenges faced in extracting subcategorization information of verbs are explored.

2.1 Urdu verb types

Verbs in Urdu can basically be divided into three types i.e., simple predicates, complex predicates and even predicates. A brief description of such classes of verbs is given in the following subsections.

2.1.1 Simple predicates

Simple predicates in Urdu comprise a single lexical element each. Different verb stems for intransitive, transitive, direct causative and indirect causative are often derived from the same base form. So, the number of such stems from a single base form can vary from one to four. Some of the base forms and different stems derived from them are listed in Table 2.1.

The phenomenon of causativization in Urdu/Hindi has been analyzed by many linguists (see, for example, Bhatt 2003; Butt 2005b; Kachru 1976; Rizvi 2008; Saksena 1982). Here, for causative stems, I mean those verbs involving an agent and one or two animated causeees (intermediary agents). However, no animate causee is involved in intransitive and transitive verbs. For example, the verb *baha* ‘make something flow’

Base Form	Intrans. Stem	Trans. Stem	Caus. Stem(s)
ʊkta	ʊkta ‘become bored’	-	-
ja	ja ‘go’	-	-
mʊrjʰa	mʊrjʰa ‘wither’	-	-
mʊskɑra	mʊskɑra ‘smile’	-	-
sɑsta	sɑsta ‘take rest’	-	-
kãp	kãp ‘shiever’	-	-
kʊrah	kʊrah ‘moan’	-	-
bʰa	-	bʰa ‘like’	-
pa	-	pa ‘find’	-
jan	-	jan ‘know’	-
pʊkar	-	pʊkar ‘call’	-
cah	-	cah ‘want’	-
xɑrid	-	xɑrid ‘buy’	-
ɑlap	-	ɑlap ‘tune up’	-
ga	ga ‘sing’	ga	-
bɑh	bɑh ‘flow’	bɑha	-
dɔṛ	dɔṛ ‘run’	-	dɔṛa
bɛṭʰ	bɛṭʰ ‘sit’	-	bɪṭʰa
cɑba	-	cɑba ‘chew’	cɑbva
kʰa	-	kʰa ‘eat’	kʰɪla/kʰɪlva
pi	-	pi ‘drink’	pɪla/pɪlva
pɑṛʰ	-	pɑṛʰ ‘read’	pɑṛʰa/pɑṛʰva
lɪkʰ	-	lɪkʰ ‘write’	lɪkʰa/lɪkʰva
pɑkɑṛ	-	pɑkɑṛ ‘hold	pɑkṛa/pɑkṛva
sʊn	-	sʊn ‘hear’	sʊna/sʊnva
bɑj	bɑj ‘be rung’	bɑja	bɑjva
bʊjʰ	bʊjʰ ‘be put out’	bʊjʰa	bʊjʰva
pɑl	pɑl ‘be brought-up’	pal	pɑlva
pɪs	pɪs ‘be crushed’	pis	pɪsva
pɪṭ	pɪṭ ‘be beaten’	piṭ	pɪṭva
dɪkʰ	dɪkʰ ‘be seen’	dekʰ	dɪkʰa/dɪkʰva
kɑṭ	kɑṭ ‘be cut’	kaṭ	kɑṭa/kɑṭva

Table 2.1: Base forms of some verbs and stems derived from them

cannot license any causee (intermediary agent), so it is a transitive stem rather than the causative stem.

2.1.1.1 Paradigms of base form derivation

Most of the base forms of simple predicates in Urdu are native roots. A very few stems of verbs have been taken from the present/past stems of Persian verbs. These are listed in Table 2.2.

Persian Infinitive	Persian Stems	Urdu Verb
azmudɑn 'test'	azmud/azma	azma
bɑxʃidɑn 'grant'	bɑxʃid/bɑxʃ	bɑxʃ
xɑridɑn 'purchase'	xɑrid/xɑr	xɑrid
ʃɑrmidɑn 'be shy'	ʃɑrmid/ʃɑrma	ʃɑrma
fɑrmudɑn 'say'	fɑrmud/fɑrma	fɑrma
gɑrdanidɑn 'regard as'	gɑrdanid/gɑrdan	gɑrdan
gʊzɑʃtɑn 'pass (by)'	gʊzɑʃt/gʊzɑr	gʊzɑr
gʊzaʃtɑn 'make pass (by)'	gʊzaʃt/gʊzar	gʊzar
nɑvaxtɑn 'caress'	nɑvaxt/nɑvaz	nɑvaz
tɑraʃidɑn 'scrape'	tɑraʃid/tɑraʃ	tɑraʃ

Table 2.2: Urdu simple verbs derived from Persian verbs

Some of the Urdu verbs have been derived from adjectives and nouns. For example the verbs *garma* 'make hot' and *narma* 'soften' have been made by adding the suffix -a to the adjectives *garm* 'hot' and *narm* 'soft' respectively. Nominal verbs are derived in many ways. Some verbs are the result of zero-derivation, that is nouns in their bare form act as verbal stems, as shown in Table 2.3.

Some nouns act as verbal stems after adding some suffix, as shown in Table 2.4. Morphologists usually claim that a noun takes a zero suffix to make a verb and then a suffix is added to causativize it. However, some intransitive verbs listed in Table 2.4 are formed by adding a suffix and in no sense can be considered entailing causativization. For example, the verb *cakra* 'feel giddy' is derived from the noun *cakar* 'circle' and it does not entail causativization.

In addition to these patterns of derivation, there are a few verbs which have been derived ideosyncratically. For example, the verb *apna* 'adopt' has been derived from the reflexive pronoun *apna* and another verb *saṭʰya* 'become mad' seems plausibly derived from the number *saṭʰ* 'sixty'. As at the age of sixty, senses of the majority of people weaken and they cannot think and react properly. This verb is used when someone even in the early age behaves like the people of sixty years age.

ɑlap ‘tuning up/to tune up’	bʰul ‘mistake/to forget’
pʊkar ‘shouting/to call’	pʰɑʈkar ‘curse/to curse’
pʰɑl ‘fruit/to grow’	pʰul ‘flower/to flourish’
tʰuk ‘saliva/to spit’	ʈʰɑg ‘scoundrel/to deceive’
cɑmɑk ‘shine/to shine’	cɑhɑk ‘chirping/to chirp’
cı̃gʰaɽ ‘scream/to scream’	cix ‘cry/to cry’
xɑrc ‘expenditure/to expend’	daɣ ‘spot/to spot’
dɑmɑk ‘shine/to shine’	dʊkʰ ‘suffering/to suffer’
dʰaɽ ‘cry/to cry’	dʰʊtkar ‘disregard/to disregard’
ɖɑ̃t ‘scolding/to scold’	ɖɑr ‘fear/to fear’
ɖıkar ‘/to devour’	rɑ̃g ‘colour/to colour’
sɑmɑjʰ ‘understanding/to understand’	kʰoj ‘search/to search’
kʰel ‘play/to play’	gɑrɑj ‘thunder/to thunder’
lɑlkar ‘shouting/to shout’	mɑhık ‘fragrance/to give fragrance’

Table 2.3: Lexemes in Urdu used both as nouns and verbs

Noun	Affix	Verbal Stem
bɑrɑf ‘ice’	a	bɑrfa ‘to cool’
fılɑm ‘film’	a	fılma ‘to cast in a film’
lalɑc ‘greed’	a	lɑlca ‘to be greedy’
laj ‘pudency’	a	lɑja ‘to be shy’
dɑfɑn ‘burial’	a	dɑfna ‘to bury’
cɑkɑr ‘circle’	a	cɑkra ‘to feel giddy’
jʰıl mıl ‘twinkling’	a	jʰılmıla ‘to twinkle’
jɑg mɑg ‘twinkling’	a	jɑgmɑga ‘to twinkle’
dɑg mɑg ‘swerving’	a	dɑgmɑga ‘to swerve’
kʰɑʈ kʰɑʈ ‘sound of knocking’	a	kʰɑʈkʰɑʈa ‘to knock’
cʰɑn cʰɑn ‘sound of bangles’	a	cʰɑncʰɑna ‘to ring’
mın mın ‘sound of bleating’	a	mınmına ‘to bleat’
tʰɑr tʰɑr ‘oscillation’	a	tʰɑrtʰɑra ‘to oscillate’
ʈɑr ʈɑr ‘croaking’	a	ʈɑrʈɑra ‘to croak’
ʈın ʈın ‘ringing’	a	ʈınʈına ‘to ring’
baɽ baɽ ‘grumbling’	a	baɽ baɽa ‘to grumble’
hın hın ‘neighing’	a	hınhına ‘to neigh’
hatʰ ‘hand’	ya	hɑtʰya ‘to grab’
jʰuʈ ‘lie’	la	jʰʊʈla ‘to falsify’

Table 2.4: Urdu simple verbs derived from nouns by adding a suffix

2.1.1.2 Paradigms of stems' derivation

Verbal roots act either as intransitive or transitive stems. More stems from base form stems are derived usually by adding suffixes. I would call the stems derived from base form stems as higher stems. The higher stems can be transitive and causatives. In Urdu, there are usually two forms of causative stems (Mohanan, 1994). One form is used when there is only one causee (1a) and the other form is used when two causees are invloved (1b). Sometimes the same form is used in both cases.

(1) a. علی نے ندا کو مسالہ چکھایا.
ɑli=ne nɪda=ko mɑsalɑh cɑk^{h}-a-ya
Ali.M.3Sg=Erg Nida.F.3Sg=Dat spices.M taste-Caus-Perf.M.3Sg
'Ali made Nida taste spices.' (Direct causative)

b. علی نے احمد سے ندا کو مسالہ چکھوایا.
ɑli=ne ɑhmɑd=se nɪda=ko mɑsalɑh
Ali.M.3Sg=Erg Ahmad.M.3Sg=Inst Nida.F.3Sg=Dat spices.M
cɑk^{h}-va-ya
taste-Caus-Perf.M.3Sg
'Ali had Ahmad make Nida taste spices.' (Indirect causative)

The base form of a verb in Urdu itself acts as one stem and other stems are derived by the inflection of the base form in different ways. The different paradigms of inflecting a single syllable base form to derive other higher stems are described in the following. The organization of paradigms, I present is a little bit different from what is made in (Chakrabarti *et al.*, 2002).

Paradigm 1: In most of cases, higher stems from the base forms are made by just adding the suffixes -*a* and -*va* without changing the vowel of the base form. Examples are given in Table 2.5.

The pattern constant-vowel-consonant (CVC) of the base form in this paradigm changes to CVCV and CVCCV in higher stems. The general syllable structure for these patterns is shown in Figure 2.1, where O, N and C stands for onset, nucleus and coda respectively.

Paradigm 2: This paradigm of derivation involves vowel strengthening. To derive the first higher stem, the short vowel of the base form is changed to the long vowel (sometimes nasalized) and the suffix -*a* is added. However the second higher stem is made by just adding the suffix -*a* or -*va* to the base form. Examples are given in Table 2.6. It should be noted that here all vowels except *ɑ*, *ɪ* and *ʊ* are considered long. The syllable structure for stems of this paradigm is shown in Figure 2.2.

Base Form	Higher Stems	
Intransitive Stem	Higher Stem I	Higher Stem II
bɑṛʰ ‘become big’	bɑṛʰ-a	bɑṛʰ-va
jɑl ‘burn’	jɑl-a	jɑl-va
cɑl ‘walk’	cɑl-a	cɑl-va
ḍɑr ‘fear’	ḍɑr-a	ḍɑr-va
kʰɪl ‘sprout’	kʰɪl-a	kʰɪl-va
gɪr ‘fall’	gɪr-a	gɪr-va
mɪṭ ‘be erased’	mɪṭ-a	mɪṭ-va
hɪl ‘shake’	hɪl-a	hɪl-va
jʰʊk ‘bend’	jʰʊk-a	jʰʊk-va
cʰʊp ‘hide’	cʰʊp-a	cʰʊp-va
dʊkʰ ‘suffer’	dʊkʰ-a	dʊkʰ-va
gʊm ‘lose’	gʊm-a	gʊm-va
Transitive Stem	Higher Stem I	Higher Stem II
pɑṛʰ ‘read’	pɑṛʰ-a	pɑṛʰ-va
lɪkʰ ‘write’	lɪkʰ-a	lɪkʰ-va
sʊn ‘listen’	sʊn-a	sʊn-va

Table 2.5: First paradigm of higher stems derivation

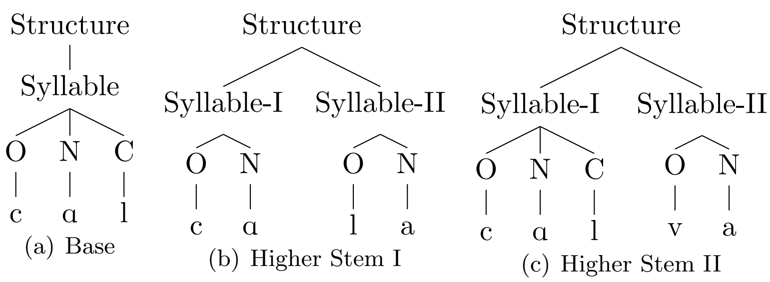


Figure 2.1: Syllable structure of Paradigm I stems

Paradigm 3: The long vowel of the base form is changed to the short vowel and the suffixes -*a* and -*va* are added. Examples are given in Table 2.7.

Paradigm 4: The long vowel of the base form is changed to the short vowel (not necessarily the corresponding one) and the suffixes -*la* and -*lva* are added. Examples are given in Table 2.8.

Paradigm 5: The coda is changed to the flap voiced retroflex. For the first stem, only the short vowel is changed to the long one and no affix is added. For the second stem, however, vowel is not changed and the suffx -*va* is added. Examples are given in Table 2.9.

Base Form	Higher Stems	
Intransitive Stem	Higher Stem I	Higher Stem II
bɑṭ 'be divided'	bɑ̃ṭ	bɑṭ-va
kɑṭ 'cut'	kaṭ	kɑṭ-a/kɑṭ-va
mɑr 'die'	mar	mɑr-a/mɑr-va
pɪṭ 'be beaten'	piṭ	pɪṭ-va
dɪk^h 'be seen'	dek^h	dɪk^h-a/dɪk^h-va
k^{h}ɪc 'be stretched'	k^{h}ɛ̃c	k^{h}ɪc-va
k^{h}ʊl 'be opened'	k^{h}ol	k^{h}ʊl-va
mʊṛ 'bend'	moṛ	mʊṛ-va
g^{h}ʊl 'melt'	g^{h}ol	g^{h}ʊl-va

Table 2.6: Second paradigm of higher stems derivation

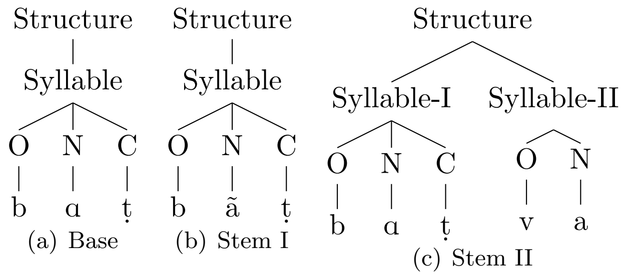


Figure 2.2: Syllable structure of Paradigm II stems

Base Form	Higher Stems	
Intransitive Stem	Higher Stem I	Higher Stem II
jag 'be awake'	jɑg-a	jɑg-va
nac 'dance'	nɑc-a	nɑc-va
bɛṭ^h 'sit'	bɪṭ^h-a	bɪṭ^h-va
leṭ 'lie'	lɪṭ-a	lɪṭ-va
p^{h}ul 'expand'	p^{h}ʊl-a	p^{h}ʊl-va
suj 'swell'	sʊj-a	sʊj-va
Transitive Stem	Higher Stem I	Higher Stem II
caṭ 'lick'	cɑṭ-a	cɑṭ-va
jit 'win'	jɪt-a	jɪt-va
b^{h}ul 'forget'	b^{h}ʊl-a	b^{h}ʊl-va

Table 2.7: Third paradigm of higher stems derivation

Base Form	Higher Stems	
Intransitive Stem	Higher Stem I	Higher Stem II
ji ‘live’	jɪ-la	jɪ-lva
ro ‘weep’	rʊ-la	rʊ-lva
so ‘sleep’	sʊ-la	sʊ-lva
Transitive Stem	Higher Stem I	Higher Stem II
kʰa ‘eat’	kʰɪ-la	kʰɪ-lva
pi ‘drink’	pɪ-la	pɪ-lva
dʰo ‘wash’	dʰʊ-la	dʰʊ-lva

Table 2.8: Fourth paradigm of higher stems derivation

Base Form	Higher Stems	
Intransitive Stem	Higher Stem I	Higher Stem II
pʰɑṭ ‘tear’	pʰaṛ	pʰɑṛ-va
pʰuṭ ‘breach’	pʰoṛ	pʰʊṛ-va
ṭuṭ ‘break’	toṛ	tʊṛ-va

Table 2.9: Fifth paradigm of higher stems derivation

In addition to these paradigms, there are further irregular patterns of derivation in the sense that they are not productive. For example, the transitive stems *cʊbʰo* ‘pinch’ and *dʊbo* ‘drown’ have been derived irregularly from their base forms *cʊbʰ* and *dub*. Another example is the verb *bec* ‘sell’ which is the transitive version of the verb *bɪk* ‘be sold’.

In this section it was shown that different base form stems are inflected in different paradigms to derive higher stems in Urdu. Suffixes, usually -*a*, -*va*, are added to both intransitive and transitive base forms for generating other transitive and/or causative stems.

2.1.2 Complex predicates

A predicate consisting of more than one predicate but giving rise to only one clausal structure is called a complex predicate (Butt, 1995; Mohanan, 1994). Complex predicates are abundantly found in South Asian languages. The first predicate usually contributes for the core meaning of the predicate. This can be a noun, a verb, an adjective. The second predicate is usually called the light verb and inflects for tense, gender and number in a sentential clause (Butt *et al.*, 1999). The most frequent light

verbs in Urdu are *kar* 'do' *ho* 'be/become', *de* 'give', *le* 'take', *a* 'come' *ja* 'go'. A variety of complex predicates found in Urdu are described in following subsections.

2.1.2.1 Verb-Verb predicates

In verb-verb complex predicates, two verbal stems are used. Only the second verbal stem inflects for agreement and the first verbal stem remains bare (2). Some examples are shown in Table 2.10.

Verb (Predicate I)	Verb (Predicate II)	Complex Predicate
caṛʰ 'climb'	dɔṛ 'run'	caṛʰ dɔṛ 'attack'
kɑr 'do'	gʊzɑr 'pass'	kɑr gʊzɑr 'do'
kɑr 'do'	de 'give'	kɑr de 'do'
kɑr 'do'	le 'take'	kɑr le 'do'
pɑṛʰ 'read'	le 'take'	pɑṛʰ le 'read'
kɑṭ 'cut'	mɑr 'die'	kɑṭ mɑr 'die'
mɑr 'die'	mɪṭ 'delete'	mɑr mɪṭ 'die'
mɑr 'die'	ja 'go'	mɑr ja 'die'
gʰɑbra 'become nervous'	ja 'go'	gʰɑbra ja 'become nervous'

Table 2.10: Verb-Verb complex predicates in Urdu

(2) a. علی نے سبق پڑھا ہے.

ɑli=ne sɑbɑq pɑṛʰ-a hɛ
Ali.M.3Sg=Erg lesson.M.3Sg read-Perf.M.3Sg be.Pres.Sg
'Ali has read the lesson.'

b. علی نے سبق پڑھ لیا ہے.

ɑli=ne sɑbɑq pɑṛʰ lɪ-ya hɛ
Ali.M.3Sg=Erg lesson.M.3Sg read take-Perf.M.3Sg be.Pres.Sg
'Ali has read the lesson.'

Sometimes English verbal stems are used with light verbs of Urdu to form complex predicates, for example, *edit kar* 'edit', *type kar* 'type', *allot kar* 'allot', *smuggle kar*

'smuggle' etc. These cannot be called as N-V complex predicates as the first predicates are not used as nouns neither in English nor in Urdu. Likewise some Arabic verbal stems are also used in complex predication that otherwise are not used, for example *qata kar* 'cut', *mana kar* 'forbid', etc.

(3) a. علی نے ندا کو پھول مارا.

ɑli=ne	nɪda=ko	pʰul	mar-a
Ali.M.3Sg=Erg	Nida.F.3Sg=Dat	flower.M.3Sg	hit-Perf.M.3Sg

'Ali hit Nida with a flower.'

b. علی نے ندا کو پھول مار دیا.

ɑli=ne	nɪda=ko	pʰul	mar	dɪ-ya
Ali.M.3Sg=Erg	Nida.F.3Sg=Dat	flower.M.3Sg	hit	give-Perf.M.3Sg

'Ali hit Nida with a flower.'

c. علی نے ندا کو پھول دے مارا.

ɑli=ne	nɪda=ko	pʰul	de	mar-a
Ali.M.3Sg=Erg	Nida.F.3Sg=Dat	flower.M.3Sg	give	hit-Perf.M.3Sg

'Ali hit Nida with a flower.'

In a very few verb-verb complex predicates, for example, the *mar de* 'hit', the order of two predicates can be reversed as in (3c). However, the verb at the second position always inflects for tense, number and gender. Poornima & Koenig (2008) have called patterns like in (3c) reverse complex predicates. Other verbal constructions in Urdu that have been analyzed as complex predicates are permissives (Butt, 1993) and some modal constructions.

(4) a. ایک پھول ندا کو سر پہ لگا.

ek	pʰul	nɪda=ko	sɑr=pɪh	lɑg-a
one	flower.M	Nida.F=Dat	head.M=Loc.on	hit-Perf.M.3Sg

'A flower struck Nida on her head.'

b. ایک پھول ندا کو سر پہ آ لگا.

ek p^{h}ul nɪda=ko sɑr=pɪh a lɑg-a
one flower.M Nida.F=Dat head.M=Loc.on come hit-Perf.M.3Sg
'A flower came and struck Nida on her head.'

c. ایک پھول ندا کو سر پہ آ کر لگا.

ek p^{h}ul nɪda=ko sɑr=pɪh a kɑr lɑg-a
one flower.M Nida.F=Dat head.M=Loc.on come Conj hit-Perf.M.3Sg
'A flower came and struck Nida on her head.'

There are some verbal combinations used in Urdu that seem to encode sequential events in a single clause like *le ʊɽ* 'take and fly', *a lag* 'come and hit', *ja lag* 'go and hit' and *kaṭ k^{h}a* 'cut and eat', etc. The subordinationg conjunction *kar/ke* can optionally be put between two morphemes of such predicates. The example *a lag* is illustrated in (4) without *kar* (4b) and with *kar* (4c). The verbal combinations of the last type have not been discussed and claimed as complex predicates in the literature.

2.1.2.2 Adjective-Verb predicates

Many subjective and objective participles from Arabic are used as adjectives in Urdu. Many such adjectives with light verbs form complex predicates. Some examples are listed in Table 2.11.

Adjective (Predicate I)	Verb (Predicate II)	Complex Predicate
ajɪz 'humble'	kɑr 'do'	ajɪz kɑr 'defeat'
hasɪl 'possessed'	kɑr 'do'	hasɪl kɑr 'get'
ʃamɪl 'included'	kɑr 'do'	ʃamɪl kɑr 'include'
raɪj 'in effect'	kɑr 'do'	raɪj kɑr 'make in effect'
maɪl 'attracted'	kɑr 'do'	maɪl kɑr 'attract'
mɑʃhur 'famous'	kɑr 'do'	mɑʃhur kɑr 'make famous'
mɑjbur 'forced'	kɑr 'do'	mɑjbur kɑr 'compel'
mɑlum 'known'	kɑr 'do'	mɑlum kɑr 'get to know'
mɑsrur 'pleased'	kɑr 'do'	mɑsrur kɑr 'please'

Table 2.11: Adjective-Verb complex predicates in Urdu

Some of the Arabic participles are only used in complex predicates in Urdu and not otherwise, for example, *ʃaɪ* is used only in the complex predicates *ʃaɪ kar* 'publish' and *ʃaɪ ho* 'be published'.

2.1.2.3 Noun-Verb predicates

There are many ways of forming a noun-verb complex predicate in Urdu. One way is to take some verbal stems which are also used as nouns in Urdu to form complex predicates with light verbs. In another combination light verbs are added to some nouns already derived from verbal stems. Some non-verbal nouns are also used in noun-verb complex predication. Examples are given in Table 2.12.

The forms of the first four nouns *ʃaram* 'shy', *bɪyah* 'marriage', *pahcan* 'recognition' and *nɪbah* 'compromise' are also used as verbs. The nouns *sʊnai* 'hearing', *dɪk*^{h}*ai* 'appearing' and *caṛ*^{h}*ai* 'climbing' are derived by adding the suffix -*ai* to verbal stems *sʊn* 'listen', *dɪk*^{h} 'appear' and *caṛ*^{h} 'climb' respectively. The noun *mɪlavaṭ* 'mixing' is derived from the verb *mɪla* 'mix' by adding the suffix -*vaṭ* and the noun *bacao* 'protection' is derived from the verb *baca* 'protect'. The noun *gʊzarɪʃ* 'request' is of Persian origin and is derived from the verb *gʊzaʃtan* 'pass' and the noun *hɪjrat* 'migration' and *nazar* 'sight' are deverbal nouns taken from Arabic. The last three nouns *nazar* 'sight', *dard* 'pain', *b*^{h}*arosah* 'trust' and *pɪyar* 'love' are not derived from verbs.

Noun (Predicate I)	Verb (Predicate II)	Complex Predicate
ʃaram 'shy'	a 'come'	ʃaram a 'to be shy'
bɪyah 'marriage'	kar 'do'	bɪyah kar 'marry'
pahcan 'recognition'	kar 'do'	pahcan kar 'recognize'
nɪbah 'compromise'	kar 'do'	nɪbah kar 'compromise'
sʊnai 'hearing'	de 'give'	sʊnai de 'hear'
dɪk^{h}ai 'appearing'	de 'give'	dɪk^{h}ai de 'be-seen'
caṛ^{h}ai 'climbing'	kar 'do'	caṛ^{h}ai kar 'compromise'
mɪlavaṭ 'mixing'	kar 'do'	mɪlavaṭ kar 'mix'
bacao 'protection'	kar 'do'	bacao kar 'protect'
gʊzarɪʃ 'request'	kar 'do'	gʊzarɪʃ kar 'request'
hɪjrat 'migration'	kar 'do'	hɪjrat kar 'migrate'
nazar 'sight'	a 'come'	nazar a 'appear'
dard 'pain'	kar 'do'	dard kar 'pain'
b^{h}arosah 'trust'	kar 'do'	b^{h}arosah kar 'trust'
pɪyar 'love'	kar 'do'	pɪyar kar 'love'

Table 2.12: Noun-Verb complex predicates in Urdu

Sometimes English deverbal nouns are also used in light verb constructions in daily conversation, e.g. *acting kar* 'act', *jogging kar* 'jog', *smoking kar* 'smoke', *swimming kar* 'swim', etc.

2.1.2.4 Compound-Verb predicates

There is also a huge variety of compound verb predicates in Urdu. Some compounds in such predicates are taken from verbal stems, some consist of a noun and a verbal stem. Perso-Arabic compounds made by either coordination of two nouns or by an *ezafe* construction are also used in light verb constructions. Examples are given in Table 2.13.

In addition to these patterns of light verb constructions in Urdu, there are still some other patterns. There are some adverbs in Urdu which are used with light verbs, for example, *vapas kar* 'return' and *samne a* 'result'. Some case phrases are used with light verbs, for example, *samaǰʰ mẽ a* 'understand', *ɪstemal mẽ la* 'use', *jan par ban* 'be in trouble', etc. Sometimes Persian prepositional phrases are also used with light verbs as in *dar gor kar* 'bury', *zer e bahs la* 'discuss', etc.

The complex predicates with Arabic and Persian compounds are rarely used in Hindi. The Persian and Arabic vocabulary and some constructions in Urdu distinguish it from Hindi. The two functional morphemes of Persian: *ezafe* and *o* are used in compounds of the d- and e- parts of Table 2.13. These are briefly described in the following.

Ezafe (e): The *ezafe* vowel in Persian appears generally between any two items that have some sort of connection (Ghomeshi, 1997). There is not a unanimous opinion on the origin of *ezafe*. Some linguists believe that it was the predicative 'is' in old Persian (Windfuhr 1979:58) and others believe that originally it is a relative pronoun (Moyne, 1971).

In Urdu, the *ezafe* construction is sometimes used in noun phrases to connect adjectival and nominal modifiers to the head noun. However, this construction is used only with the Persian loan words. The orthographic form of *ezafe* in the Urdu script differs in different contexts. In contrast with the genitival modifiers in Urdu which appear usually in the prenominal position, the modifiers in the *ezafe* construction appear after the head (Bögel & Butt, to appear). It is illustrated in (5).

Compound	Verb	Complex Predicate
a. Compounds made from two verbal stems		
ʊkʰaṛ pɑcʰaṛ ‘felling’	kɑr ‘do’	ʊkʰaṛ pɑcʰaṛ kɑr ‘fell’
ʊcʰɑl kud ‘playing’	kɑr ‘do’	ʊcʰɑl kud kɑr ‘play’
cir pʰaṛ ‘tearing’	kɑr ‘do’	cir pʰaṛ kɑr ‘tear’
qɑlɑ qɑmɑ ‘crushing’ (Arabic)	kɑr ‘do’	qɑlɑ qɑmɑ kɑr ‘crush’
tɑraʃ xɑraʃ ‘sharpening’ (Persian)	kɑr ‘do’	tɑraʃ xɑraʃ kɑr ‘sharpen’
b. Compounds made from a noun and an inflected verbal stem		
bix kɑni ‘uprooting’ (Persian)	kɑr ‘do’	bix kɑni kɑr ‘uproot’
caʃm poʃi ‘connivance’ (Persian)	kɑr ‘do’	caʃm poʃi kɑr ‘connive’
sɪgret noʃi ‘smoking’ (Persian)	kɑr ‘do’	sɪgret noʃi kɑr ‘smoke’
c. Compounds made from a noun and a verbal stem		
ɑsɑr ɑndaz ‘effective’ (Persian)	ho ‘be/become’	ɑsɑr ɑndaz ho ‘affect’
nɑzɑr ɑndaz ‘ignored’ (Persian)	kɑr ‘do’	nɑzɑr ɑndaz kɑr ‘ignore’
sɑf ara ‘being in lines’ (Persian)	ho ‘be/become’	sɑf ara ho ‘become ready’
d. Ezafe compounds (nouns)		
sɑrf e nɑzɑr ‘ignoring’ (Persian)	kɑr ‘do’	sɑrf e nɑzɑr kɑr ‘ignore’
nɑzr e kɑrɑm ‘pity’ (Persian)	kɑr ‘do’	nɑzr e kɑrɑm kɑr ‘pity’
ɣɑrq e ab ‘immersion’ (Persian)	ho ‘be/become’	ɣɑrq e ab ho ‘drown’
e. Coordination compounds (nouns)		
jɪd o jʊhd ‘effort’ (Arabic)	kɑr ‘do’	jɪd o jʊhd kɑr ‘make effort’
tɑg o do ‘effort’ (Persian)	kɑr ‘do’	tɑg o do kɑr ‘do’ ‘try’
qɑtl o ɣɑrɑt ‘massacre’ (Arabic)	kɑr ‘do’	qɑtl o ɣɑrɑt kɑr ‘massacre’
kʊʃt o xun ‘massacre’ (Persian)	kɑr ‘do’	kʊʃt o xun kɑr ‘massacre’
f. Other compounds (nouns)		
bat cit ‘conversation’	kɑr ‘do’	bat cit kɑr ‘talk’
ʃor ʃɑrabɑh ‘noise’	kɑr ‘do’	ʃor ʃɑrabɑh kɑr ‘scream’
kam kaj ‘work’	kɑr ‘do’	kam kaj kɑr ‘work’
tɪkkɑh boṭi ‘piece’	kɑr ‘do’	tɪkkɑh boṭi kɑr ‘cut’

Table 2.13: Compound-Verb complex predicates in Urdu

(5) a. پاکستان کی حکومت

pakɪstan=ki hɑkumɑt
Pakistan=Gen.F government.F
‘Government of Pakistan’

b. حکومت پاکستان

hɑkumɑt e pakıstan
government.F Ez Paskistan
'Government of Pakistan'

(6) a. دل میں ذوق وصل و یاد یار تک باقی نہیں

آگ اس گھر میں لگی ایسی کہ جو تھا جل گیا

dıl=mẽ [zɔq e vɑsl] o [yad e yar] tɑk baqi nɑhĩ
heart=in verve Ez union and memory Ez beloved even left not

ag ıs gʰɑr=mẽ lɑg-i ɛsi kıh jo tʰa jɑl gɑ-ya
fire this house=in set-Perf such that what was burn go-Perf

'There is not left the verve of the union and the memory of the beloved; such a fire set to this house that every thing burnt' (Asadullah Ghalib)

b. تیشے بغیر مر نہ سکا کوہکن اسد

سرگشتہ خمار رسوم و قیود تھا

teʃe bɑɣɛr mɑr nɑh sɑk-a koh-kɑn ɑsɑd
adze.M without die not can-Perf.M mountain-digger Asad

sɑrgɑʃtɑh e xʊmar e [rɑsum o qʊyud] tʰa
bewildered Ez wine-headache Ez tradition.3Pl and limit.3Pl be.Past

'The mountain digger (Farhad) could not die without adze; he was bewildered of the headache of traditions and limits.' (Asadullah Ghalib)

Conjunction *o*: The form 'و' is used both in Arabic and in Persian as a coordinating conjunction in the meaning of 'and'. It is sometimes pronounced as *vao* and sometimes just as *o*. This conjunction form is used in compounds of the e-part of Table 2.13.

Like *ezafe*, the conjunction form *o* 'and' in Urdu is also used productively with Arabic and Persian loan words, but not with purely native words. In Urdu, the noun phrases in an *ezafe* construction can also be conjoined using this conjunction form (6a) and also the noun phrases involving this conjunction can be used in an *ezafe*

construction (6b). This form, however, is never used to conjoin sentences in Urdu in contrast with Arabic and Persian. The equivalent native coordinating conjunction form of Urdu is *ɔr* ‘and’ which is also used for conjoining clauses.

In complex predicates, two predicates are not always contiguous. For example, the negative element *nahĩ* ‘not’ can appear between the main verb and the light verb in a complex predicate.

2.1.3 Even predicates

In many Indo-Aryan languages, sometimes two semantically related verbs are used together in a more abstract meaning that might not be exactly the same as encoded by individual verbs. Such verbs, I call as even predicates. Typical examples are listed in Table 2.14.

Predicate I	Predicate II	Even Predicate
a ‘come’	ja ‘go’	a ja ‘keep visiting’
pɑṛ^h ‘read’	lɪk^h ‘write’	pɑṛ^h lɪk^h ‘get education’
p^{h}ɑl ‘bear fruit’	p^{h}ʊl ‘bear flowers’	p^{h}ɑl p^{h}ʊl ‘grow’
jãc ‘test’	pɑrɑk^h ‘test’	jãc pɑrɑk^h ‘test’
cɑl ‘walk’	p^{h}ɪr ‘revolve’	cɑl p^{h}ɪr ‘wander’
g^{h}um ‘rotate’	p^{h}ɪr ‘revolve’	g^{h}um p^{h}ɪr ‘wander’
cir ‘tear’	p^{h}aṛ ‘tear’	cir p^{h}aṛ ‘tear’
ro ‘weep’	d^{h}o ‘wash’	ro d^{h}o ‘weep’
lɑṛ ‘quarrel’	j^{h}ɑgɑṛ ‘quarrel’	lɑṛ j^{h}ɑgɑṛ ‘quarrel’
k^{h}a ‘eat’	pi ‘drink’	k^{h}a pi ‘eat and drink’
k^{h}el ‘play’	kud ‘jump’	k^{h}el kud ‘play’
ʊc^{h}al ‘jump’	kud ‘jump’	ʊc^{h}al kud ‘play’
mɪl ‘meet’	jʊl ‘meet’	mɪl jʊl ‘meet’
g^{h}ʊl ‘melt’	mɪl ‘meet’	g^{h}ʊl mɪl ‘become frank’
nac ‘dance’	ga ‘sing’	nac ga ‘make merry’
soc ‘think’	sɑmɑj^h ‘understand’	soc sɑmɑj^h ‘ponder’
dek^h ‘see’	b^{h}al ‘see’	dek^h b^{h}al ‘investigate’
lɑga ‘ignite’	bʊj^{h}a ‘put out’	lɑga bʊj^{h}a ‘infuriate’
pɑr^{h}a ‘teach’	sɪk^{h}a ‘teach’	pɑr^{h}a sɪk^{h}a ‘educate/train’
bɑr^{h}a ‘increase’	cɑr^{h}a ‘cover’	bɑr^{h}a cɑr^{h}a ‘exaggerate’

Table 2.14: Even predicates in Urdu

Both verbs in even predicates are inflected similarly for tense, number and gender as is shown in example sentences in (7). So, we cannot distinguish two verbs in

even predicates as the main verb and the light verb because the meanings of the even predicates are contributed by both verbs and both are inflected the same in a clause.

(7) a. سلسلے توڑ گیا وہ سبھی جاتے جاتے
اتنے تو مراسم تھے کہ آتے جاتے

sılsıle	toṛ	gɑ-ya	vʊh	sɑbʰi	ja-te	ja-te
chain.M.3Pl	break	go-Perf	3Sg	all	go-Imperf	go-Imerf

ıtne	to	mɑrasım	tʰe	kıh	a-te	ja-te
such	Emp	relationship.M.3Pl	be.3Pl.Past	that	come-Perf	go-Perf

‘He broke all chains while leaving; there was at least such a relationship, that he should have keep visiting’ (Ahmad Faraz)

b. علی ندا کے ہاں آتا جاتا ہے.

ɑli	nıda=ke	hã	a-ta	ja-ta	hɛ
Ali.M	Nida.F=Gen	side	come-Imperf.M.3Sg	go-Imperf.M.3Sg	be.Pres.Sg

‘Ali keeps visiting Nida.’

c. بچے تفریح میں خوب کھیلے کودے.

bɑcce	tɑfrih=mẽ	xub	kʰel-e	kud-e
child.Pl	recess.F=Loc.in	well	play-Perf.Pl	jump-Perf.Pl

‘The children played well in the recess.’

In (7a) the phrase *ja-te ja-te* is an alternate construction of *ja-te hu-e* which means ‘while leaving’ and the other phrase *a-te ja-te* ‘keep visiting’ due to bare imperfect aspect gives the subjunctive meaning entailing wish. Another simple example of this even predicate is given in (7b) and in (7c), the use of another even predicate *kʰel kud* ‘play’ is illustrated.

2.1.4 Summary

In this section, different types of verbs in Urdu have been described. There are basically three types of verbs i.e. simple predicates, complex predicates and even predicates. In

many of complex predicates, the main predicate is from Persian or Arabic vocabulary. To identify different types of complex predicates from the corpus automatically is not straightforward because the two parts of a complex predicate are not always contiguous to each other in a sentence. The scope of the work in this thesis is limited to acquire the subcategorization frames of simple predicates. A repository of simple predicates has been developed from different resources. The challenges faced with acquiring subcategorization information of simple predicates are reported in the next section.

2.2 Challenges for subcategorization acquisition

Like many other Indo-Aryan languages, Urdu is a case rich language. Different case forms mark actants of a clause for different cases. The oblique form of a noun is used when it is marked for case as it is shown in (8b) for the common noun *bacca* 'baby'. In (8a), where this noun is unmarked, its direct form is used. In addition to proper nouns, many common nouns also have the same direct and oblique forms. The scope of case markers in Urdu extends to every noun of coordination (8c). Due to their phrasal scope, these are termed as case clitics (Butt & King, 2005).

(8) a. بچہ بہت شرارتی ہے۔
bɑccɑ bɑhʊt ʃɑrarti hɛ
baby.Dir.Nom very naughty be.Pres.3Sg
'The baby is very naughty.'

b. ندا نے بچے کو ڈانٹا۔
nɪda=ne bɑce=ko dãṭ-a
Nida.F=Erg baby.Obl=Acc scold-Perf.M.3Sg
'Nida scolded the baby.'

c. علی اور ندا نے بچے کو ڈانٹا۔
[ɑli ɔr nɪda]=ne bɑce=ko dãṭ-a
Ali and Nida.F=Erg baby=Acc scold-Perf.M.3Sg
'Ali and Nida scolded the baby.'

(9) a. ندا سیب کھاتی ہے.

nıda	seb	k^ha-ti	hɛ
Nida.F	apple.M.3Sg	eat-Imperf.F.3Sg	be.Pres.3Sg

‘Nida eats an apple.’

b. ندا نے سیب کھایا.

nıda=ne	seb	k^ha-ya
Nida.F=Erg	apple.M.3Sg	eat-Perf.M.Sg

‘Nida ate an apple.’

Urdu is considered to be a morphologically split-ergative language. The case on subjects of transitive clauses alternates between ergative and nominative (9). The ergative marker is sensitive to the perfective aspect (e.g., Anderson 1977, Mahajan 1990, Dixon 1994). In addition to ergative-nominative pattern for transitive verbs, this pattern also exists for some intransitive verbs in Urdu (Butt & King, 2005).

For Urdu, being a case-rich language, it apparently seems trivial to recognize arguments of predicates based on case clitic cues but the complexity of the case system and free word order nature of the language actually make the task very difficult. The following five issues present challenges for automatic acquisition of subcategorization frames of verbs in Urdu from a raw corpus.

2.2.1 Absence of unique case clitic forms

There are several different case clitic forms: ϕ, *ne*, *ko*, *se*, *mẽ*, *par*, *tak* in Urdu. The clitic form ϕ stands for the unmarked case. There is not always a one to one correspondence between case clitic form and case function.

2.2.1.1 The case clitic *ko*

The same clitic form *ko* can mark nouns for different cases: accusative, dative, locative and temporal (Mohanan, 1994). Consider the example sentences in (10).

(10) a. علی نے ندا کو بلایا.

ɑli=ne	nɪda=ko	bʊla-ya
Ali.M.3Sg=Erg	Nida.F.3Sg=Acc	call-Perf.M.3Sg

‘Ali called Nida.’

b. علی نے ندا کو پھول دیا.

ɑli=ne	nɪda=ko	p^{h}ul	dɪ-ya
Ali.M.3Sg=Erg	Nida.F.3Sg=Dat	flower.M.3Sg	give-Perf.M.3Sg

‘Ali gave a flower to Nida.’

c. علی رات کو آیا.

ɑli	rat=ko	a-ya
Ali.M.3Sg.Nom	night.F.3Sg=Temp	come-Perf.M.3Sg

‘Ali came at night.’

d. علی گھر کو گیا.

ɑli	g^{h}ɑr=ko	gɑ-ya
Ali.M.3Sg.Nom	home.M.3Sg=Loc	go-Perf.M.3Sg

‘Ali went home.’

The direct object of a transitive verb *bʊla* ‘call’ in (10a) and the indirect object of ditransitive verb *de* ‘give’ in (10b) are marked accusative and dative respectively by the same case clitic form *ko*. Likewise the temporal and locative object in (10c–d) are marked by the same clitic form *ko*. Still more uses of the clitic *ko* (Ahmed, 2006) are given in (11)–(12).

(11) a. علی جانے کو ہے.

ɑli	ja-ne=ko	hɛ
Ali.M.3Sg	go-Inf.Obl=for	be.Pres.Sg

‘Ali is about to go.’

b. علی جانے والا ہے.

ɑli	ja-ne=vala	hɛ
Ali.M.3Sg	go-Inf.Obl=VA	be.Pres.Sg

‘Ali is about to go.’

(12) a. ندا کے پاس دینے کے لیے کچھ نہیں.

nıda=ke pas de-ne=ke lıye kʊcʰ nɑhĩ
Nida.F=Gen near give-Inf.Obl=Gen for someting not
'Nida has nothing for giving.'

b. ندا کے پاس دینے کو کچھ نہیں.

nıda=ke pas de-ne=ko kʊcʰ nɑhĩ
Nida.F=Gen near give-Inf.Obl=KO someting not
'Nida has nothing for giving.'

c. دل کی خوشی کی خاطر چکھ ڈال مال و دھن کو

گر مرد ہے تو عاشق کوڑی نہ رکھ کفن کو

dıl=ki xʊʃi=ki xatır cɑkʰ ḍal mal o dʰɑn=ko
heart=Gen.F joy.F for devour put property and wealth=Acc
gɑr mɑrd hɛ tu aʃıq koṛi nɑh rɑkʰ kɑfɑn=ko
if man be.3Sg.Pres you lover penny not put shroud=KO
'For the joy of heart, devour all of your property and wealth; If you really are a lover, leave not a penny for your shroud' (Nazir Akbar Abadi)

An immediate action clause can be constructed by adding the *ko* cilitc to the oblique form of an infinitve followed by the verb *ho* 'be' as is shown in (11a). In another construction for encoding this notion, the clitic *vala* is used (11b) with the infinitive. Sometimes the purpose notion that usually is encoded by using the postposition *lıye* 'for' (12a) is also encoded instead by using the clitic *ko* (12b–c).

2.2.1.2 The case clitic *se*

Like the clitic *ko*, the clitic *se* also has many functions (Ahmed, 2011; Mohanan, 1994), more than any other case clitic in Urdu. It is used with instrumental adjuncts as in (13a), for source expressions; humans, as in (13b), material, as in (13c), locatives, as in (13d) as well as for temporal expressions (13e).

(13) a. علی نے چاقو سے سیب کاٹا.

ɑli=ne	caqu=se	seb	kat-a
Ali.M.3Sg=Erg	knife.M.3Sg=Instr	apple.M.3Sg	cut-Perf.M.3Sg

‘Ali cut the apple with a knife.’

b. علی نے ندا سے کتاب لی.

ɑli=ne	nɪda=se	kɪtab	l-i
Ali.M.3Sg=Erg	Nida.F.3Sg=Abl	book.F.3Sg	get-Perf.F.Sg

‘Nida got a book from Nida.’

c. ووڈکا عموما چقندر سے بنتی ہے.

voḍka	ʊmumɑn	cʊqɑndɑr=se	bɑn-ti	hɛ
vodka.F.Sg	usually	beet.M=Abl	make-Imperf.F.Sg	be.Pres.Sg

‘Vodka is usually made from beets.’

d. علی ملتان سے آیا.

ɑli	mʊltan=se	a-ya
Ali.Nom.M.3Sg	Multan=Loc	come-Perf.M.3Sg

‘Ali came from Multan.’

e. ندا کل سے بیمار ہے.

nɪda	kɑl=se	bimar	hɛ
Nida.F.3Sg	yesterday=Temp	ill	be.Pres.3Sg

‘Nida is ill from yesterday.’

The clitic *se* is also used as a comitative case marker as in (14a) and to mark the causee in causative constructions as in (14b).

(14) a. علی نے ندا سے گفتگو کی.

ɑli=ne	nɪda=se	gʊftʊgu	ki
Ali.M.3Sg=Erg	Nida.F.3Sg=Com	conversation.F.3Sg	make.Perf.F.Sg

‘Ali made a conversation with Nida.’

b. علی نے ندا سے چٹھی لکھوائی.

ɑli=ne nɪda=se cɪṭṭ^{h}i lɪk^h-va-i

Ali.M.3Sg=Erg Nida.F.3Sg=Abl letter.F.3Sg write-Caus-Perf.F.3Sg

'Ali made Nida write the letter.'

The clitic *se* marks the logical subject (demoted agent) in the standard passive constructions (15a) as well as the subject in passives of disability (15b). The (dis)ability passive differs from the standard passives in that unlike the standard passives it is also possible with intransitives. Davison (1990) has called them peculiar passives, Butt (1997) has analyzed such constructions as complex predicates with dispositional semantics whereas Bhatt (1998) has analysed them in terms of negative polarity. The same clitic *se* is also used for comparative expressions (15c) and used with some nouns to give the adverbial meanings of manner or time (15d).

(15) a. چٹھی ندا سے لکھی گئی.

cɪṭṭ^{h}i nɪda=se lɪk^h-i gɑ-i

letter.F.3Sg Nida.F.3Sg=Abl write-Perf.F.3Sg go-Perf.F.3Sg

'The letter was written by Nida.'

b. علی سے چلا نہیں گیا.

ɑli=se cɑl-a nɑhĩ gɑ-ya

Ali.M.3Sg=Abl walk-Perf.M.3Sg not go-Perf.M.Sg

'Ali could not walk.'

c. ندا علی سے چھوٹی ہے.

nɪda ɑli=se c^{h}oṭi hɛ

Nida.F.3Sg Ali.Nom.M.3Sg=Abl younger be.Pres.Sg

'Nida is younger than Ali.'

d. علی گھر دیر سے آیا.

ɑli g^{h}ɑr der=se a-ya

Ali.M.3Sg home.M.3Sg late=Abl come-Perf.M.3Sg

'Ali came home late.'

(16) a. علی نے کار سے سفر کیا.

ɑli=ne kar=se sɑfɑr kɪ-ya
Ali.M=Erg car.F=Abl travel.M do-Perf.M.Sg
'Ali travelled by a car.'

b. علی نے سڑک سے سفر کیا.

ɑli=ne sɑṛɑk=se sɑfɑr kɪ-ya
Ali.M=Erg car.F=Abl road.F do-Perf.M.Sg
'Ali travelled by road.'

(17) a. مجرم گرفتاری کے ڈر سے چھپ گیا.

mʊjrɪm gɪrɪftari=ke ḍɑr=se c^{h}ʊp gɑ-ya
culprit.M arrest.F=Gen fear.M=Abl hide go-Perf.M.3Sg
'The culprit hid himself due to the fear of arrest.'

b. افسوس بے شمار سخن ہائے گفتنی
خوف فساد خلق سے ناگفتہ رہ گئے

ɑfsos be-ʃʊmar sʊxɑn-ha e gʊftɑni
alas un-countable point-Pl Ez worth-to-be-mentioned

xɔf e fɑsad e xɑlq=se na-gʊftɑh rɑh gɑe
fear Ez violence Ez people=Abl unuttered remain go-Perf.Pl

'Alas, many points worth to be mentioned remained unuttered due to the fear of violence of people' (Majid Amjad)

The marker *se* is also used to mark the means of traveling (Rizvi, 2007), as shown in (16). Sometimes the clitic *se* is used in the meaning of 'because of', as shown in (17). In addition to all above uses, the clitic *se* is also used for path (Ahmed, 2011) in Urdu as in *baɣ se* 'through the garden'.

2.2.2 Different marking of grammatical functions

We have seen in previous section that a single case form is used to mark for different case functions. Alongside this, the grammatical functions are also not marked uniquely

for case in Urdu (Butt & King, 2005; Mohanan, 1994). Different possibilities of marking subject and object are discussed in the following subsections.

2.2.2.1 Different marking of subject

The subject in Urdu is marked for many different cases in different constructions. Examples are given in (18).

(18) a. ندا جا رہی ہے.

nɪda ja rɑh-i hɛ
Nida.F.3Sg go Prog-F.3Sg be.Pres.3Sg
'Nida is leaving.'

b. ندا نے کچھ نہیں کھایا.

nɪda=ne kʊc^h nɑhĩ k^{h}a-ya
Nida.F.3Sg=Erg something not eat-Perf.M.3Sg
'Nida has eaten nothing.'

c. ندا سے چلا نہیں جاتا.

nɪda=se cɑl-a nɑhĩ ja-ta
Nida.F.3Sg=Inst walk not go-Imperf.M.3Sg
'Nida cannot walk.'

d. ندا کو گھر جانا ہے.

nɪda=ko g^{h}ɑr ja-na hɛ
Nida.F.3Sg=Dat home.M go-Inf.M.3Sg be.Pres.3Sg
'Nida has to go home.'

e. ندا کی ایک بیٹی ہے.

nɪda=ki ek beṭi hɛ
Nida.F.3Sg=Gen one daughter.F.3Sg be.Pres.3Sg
'Nida has one daughter.'

f. ندا میں بڑی مامتا ہے۔
nɪda=mẽ bɑṛi mamta hɛ
Nida.F.3Sg=Loc.in much affection be.Pres.3Sg
'Nida has much affection.'

g. ندا پر بیٹی کی پرورش کی ذمہ داری ہے۔
nɪda=pɑr beṭi=ki pɑrvɑrɪʃ=ki zɪmmɑh-dari hɛ
Nida.F=Loc.on daughter.F=Gen upbringing.F=Gen responsibility be.Pres
'Nida has the responsibility of upbringing the daughter.'

h. بیٹی کو چوما گیا۔
beṭi=ko cum-a gɑ-ya
daughter.F=Acc kiss-Perf.M.3Sg go-Perf.M.3Sg
'The daughter was kissed.'

The case of the subject in (18a-g) is nominative, ergative, instrumental, dative, genitive, in-locative and on-locative respectively. The accusative case of grammatical subject is possible in passives of transitive verbs (18h).

2.2.2.2 Differential object marking

The direct object of a verb can be unmarked or accusative (Butt & King, 2005), for example, consider the following sentences:

(19) a. علی نے سیب کھایا۔
ɑli=ne seb k^ha-ya
Ali.M.3Sg=Erg apple.M.3Sg.Nom eat-Perf.M.3Sg
'Ali ate an apple.'

b. علی نے سیب کو کھایا۔
ɑli=ne seb=ko k^ha-ya
Ali.M.3Sg=Erg apple.M.3Sg=Acc write-Perf.M.3Sg
'Ali ate the apple.'

(20) a. علی نے ندا کو سونے کا کہا.
ɑli=ne nɪda=ko so-ne=ka kɑ-ha
Ali.M.3Sg=Erg Nida.F.3Sg=Acc sleep-Inf=Gen ask-Perf.M.Sg
'Ali asked Nida to sleep.'

b. علی نے ندا سے سونے کا کہا.
ɑli=ne nɪda=se so-ne=ka kɑ-ha
Ali.M.3Sg=Erg Nida.F.3Sg=Com sleep-Inf=Gen ask-Perf.M.Sg
'Ali asked Nida to sleep.'

The direct object *seb* 'apple' in (19a) is null-marked for nominative case and it is *ko*-marked for accusative case in (19b). Specific direct objects in Urdu are marked accusative (Butt, 1993). The object of some verbs is either marked accusative (20a) or comitative (20b). In Table 2.15 clitics' forms, cases and their associations with grammatical functions are listed. This table is a slightly modified version of what is given in Butt & King (2005).

Case clitic	Case feature	Grammatical function
ϕ	Nominative	Subject, Direct object
	Locative	Locative object
	Temporal	Temporal object
ne	Ergative	Subject
ko	Accusative	Direct object
	Dative	Subject, Indirect object
	Temporal	Temporal object
	Locative	Locative object
se	Instrumental	Instrument
	Comitative	Object
	Ablative	Indirect object
	Locative	Locative object
	Ability	Subject
mẽ	Locative	Locative object
	Temporal	Temporal object
pɑr	Locative	Locative object
	Temporal	Temporal object
tɑk	Locative	Locative object
	Temporal	Temporal object

Table 2.15: Case and grammatical functions

2.2.3 Free word order

Fixed order in a language like English is useful in that the order itself provides cues for recognizing the arguments of a verb in the sentence. However, Urdu is a free word order language. The verb in a sentence usually (but not always) comes last and its arguments are put in any order before it. For example, the arguments of the verb *de* ‘to give’ in (10b) can be arranged in the different orders shown in (21). The truth-conditional meaning of the sentence remains the same, regardless of the order of the verb’s arguments.

(21) a. علی نے ندا کو پھول دیا.

ɑli=ne	nıda=ko	p^hul	dı-ya
Ali.M.3Sg=Erg	Nida.F.3Sg=Dat	flower.M.3Sg.Nom	give-Perf.M.3Sg

‘Ali gave a flower to Nida.’

b. علی نے پھول ندا کو دیا.

ɑli=ne	p^hul	nıda=ko	dı-ya
Ali.M.3Sg=Erg	flower.M.3Sg.Nom	Nida.F.3Sg=Dat	give-Perf.M.3Sg

c. ندا کو علی نے پھول دیا.

nıda=ko	ɑli=ne	p^hul	dı-ya
Nida.F.3Sg=Dat	Ali.M.3Sg=Erg	flower.M.3Sg.Nom	give-Perf.M.3Sg

d. ندا کو پھول علی نے دیا.

nıda=ko	p^hul	ɑli=ne	dı-ya
Nida.F.3Sg=Dat	flower.M.3Sg.Nom	Ali.M.3Sg=Erg	give-Perf.M.3Sg

e. پھول علی نے ندا کو دیا.

p^hul	ɑli=ne	nıda=ko	dı-ya
flower.M.3Sg.Nom	Ali.M.3Sg=Erg	Nida.F.3Sg=Dat	give-Perf.M.3Sg

f. پھول ندا کو علی نے دیا.

p^hul	nıda=ko	ɑli=ne	dı-ya
flower.M.3Sg.Nom	Nida.F.3Sg=Dat	Ali.M.3Sg=Erg	give-Perf.M.3Sg

Grammatical functions of nouns in a sentence in Urdu can be identified by case and animacy features of participants rather than their position in the sentence. The verb *de* 'to give' is a ditransitive verb and takes an ergative subject in (21). As already said, Urdu is a split ergative language. The ergative subject is taken by verbs with conditions of transitivity, tense/aspect, choice of light verb and volitionality (Butt, 2006). The subject in (21) will be marked nominative in case the aspect is not perfect (22).

(22) علی ندا کو پھول دیتا ہے۔

ɑli	nɪda=ko	pʰul	de-ta	hɛ
Ali.M.3Sg.Nom	Nida.F.3Sg=Dat	flower.M.3Sg.Nom	give-Perf.M.3Sg	be.3Sg.Pres

'Ali gives a flower to Nida.'

(23) مکان میں آگ لگ گئی۔

mɑkan=mẽ	ag	lɑg	gɑ-i
house.M.3Sg=Loc.in	fire	set	go-Perf.F.3Sg

'The house caught the fire.'

In (22), the two arguments of the predicate that is *ali* 'Ali' and *pʰul* 'flower' are nominative and the subject *ali* 'Ali' in such a case could be recognized due to its animacy feature. The *ko* marked argument can be identified as a dative marked argument only when the verb is tagged as a ditransitive verb. So, an Urdu corpus tagged for the case and animacy features of nouns and types of verbs could be helpful for the automatic recognition of the subject and other arguments of most of the verbs. However, it would be hard to distinguish arguments of those verbs which generally do not involve animate participants (23). The mere position of the argument in Urdu is not helpful for recognition of arguments in contrast with English and many other langauges.

2.2.4 Multifunctionality of the complementizer *kɪh*

Some verbs in Urdu take *that*-clause as their argument in a sentence. The complementizer form *kɪh* is used to construct the complementizer clause as shown in (24).

(24) ندا نے کہا کہ وہ کل نہیں آئے گی.
nɪda=ne kɑh-a [kɪh vʊh kɑl nɑhĩ a-e-gi]
Nida.F=Erg tell-Perf.3Sg that 3Sg.F tomorrow not come-Subjn-Fut.F
'Nida told that she would not come the next day.'

To recognize the complementizer clause argument of a verb in Urdu by identifying complementizer form *kɪh* in a sentence is also not straightforward as the same complementizer form also has some other functions: relative pronoun, conjunction, temporal. Different uses of *kɪh* are illustrated in (25).

(25) a. ہم نے مانا کہ تغافل نہ کرو گے.
hɑm=ne man-a [kɪh tɑɣafʊl nɑh kɑr-o-ge]
1.Pl=Erg agree-Perf that.Comp negligence.F.3Sg not do-Subjn.2-Fut.3Sg
'We agreed that you would not be negligent.'

b. اچھا تھا کہ ندا آتی.
ɑc^{h}a t^{h}a [kɪh nɪda a-ti]
good be.3Sg.Past that Nida.F come-Imperf.F.3Sg
'It would have been good if Nida had come.'

c. کتاب جو کہ میز پر پڑی ہے لاؤ.
kɪtab jo kɪh mez=pɑr pɑṛ-i hɛ, la-o
book.F.Nom that.RelP table.F=Loc lie-Perf.F be.Subjn bring-Imper
'Bring the book that is good.'

d. علی اول آیا یا کہ ندا؟
ɑli ɑvɑl a-ya ya/kɪh nɪda
Ali.M.3Sg.Nom first come-Perf.3Sg or Nida.F.3Sg
'Ali stood first or Nida?'

e. علی گھر سے نکلا ہی تھا کہ بارش برسنے لگی۔

ɑli	gʰɑr=se	nikl-a	hi	tʰ-a
Ali.M.3Sg	home.M.3Sg=Abl	leave-Perf.3Sg	Emp	be.Past-3Sg

kɪh	barɪʃ	bɑrɑs-ne	lɑg-i
that	rain.3Sg.F	shower-Inf	start-Perf.3Sg.F

'It started raining when Ali just left the house.'

f. خطا تو جب ہو کہ ہم حال دل کسی سے کہیں

کسی کو چاہتے رہنا کوئی خطا تو نہیں

xɑta	to	**jɑb**	ho	**kɪh**
offense=Gen	Emp	when	be.Subjn	that

hɑm	hal	e	dɪl	kɪsi=se	kɑhẽ
1Sg.Pl	condition	Ez	heart	anyone=Abl	say.Subjn.Pl

kɪsi=ko	cah-te	rɑh-na	koi	xɑta	to	nɑhĩ
anyone=Acc	want-Imperf.Pl	remain-Inf	some	offense.F	Emp	not

'Offense should only be considered if we say our condition of heart to someone; merely to like someone is not an offense at all.' (Sameer)

The form *kih* in sentence (25a) is used as a canonical complementizer, in (25b) it is used to construct an if-clause, in (25c) it is used as a relative pronoun, in (25d) as a conjunction and in (25e–f) as a temporal marker. The same form in Urdu poetry is sometimes also used in place of *balkɪh* 'but' (Khan, 2005) as in (26a) and sometimes *kɪyɔ̃kɪh* 'because' as in (26b). In addition to this, with the use of some adverbs like *ɪtna* 'this much', *ɛsa* 'such' the form *kɪh* can appear with any verb in Urdu (27).

(26) a. نہیں ملتی کوشش سے دنیا ہی تنہا

کہ ارکان دیں بھی اسی پر ہیں برپا

nɑhĩ	mɪlti	koʃɪʃ=se	dʊnya	hi	tɑnha
not	be-got.Imperf.3Sg	effort.3Sg=Abl	world.F.3Sg	Emp	only

kɪh	ɑrkan-e	dĩ	bʰi	ɪsi=par	hɛ̃	bɑrpa
but	unit.3Pl-Ez	religion	also	this=Loc.on	be.Pres.3Pl	fixed

'Not only the worldly benefits we get from effort,
but the units of the religion are also fixed on it.' (Altaf Hussain Hali)

b. تیری خاک میں ہے گر شرر تو خیال فقر و غنا نہ کر

کہ جہاں میں نان شعیر پر ہے مدار قوت حیدری

teri xak mẽ hɛ gɑr ʃɑrɑr
your.F origin.F Loc.in be.Cop if spark

to xɑyal e fɑqr o ɣɪna nɑh kɑr
then consideration Ez poor-being and well-being not do.Imper

kɪh jɑhɑ̃ mẽ nan e ʃɑ'ir pɑr
that.Compl world Loc.in bread.M Ez barley Loc.on

hɛ mɑdar e qʊvat e hɛdri
be.Cop dependence.M Ez power.F Ez of-Haider

'If there is a spark in your origin then do not think of poverty or richness; as a power the like of Haider depends only on barley bread.' (Allama Iqbal)

(27) a. ندا اتنا روئی کہ اس کی آنکھیں سرخ ہو گئیں.

nɪda **ɪtna** ro-i
Nida.F.3Sg so much weep-Perf.F.3Sg

kɪh ʊs=ki ãk^{h}ẽ sʊrx ho gɑ-ĩ
that 3Sg=Gen eye.F.3Pl red become go-Perf.F.3Pl

'Nida wept so much that her eyes became red.'

b. لاغر اتنا ہوں کہ گر تو بزم میں جا دے مجھے

میرا ذمہ دیکھ کر گر کوئی بتلا دے مجھے

laɣɪr **ɪtna** hũ **kɪh** gɑr
feeble such be.Cop.1Sg that.Compl if

tu bɑzm mẽ ja de mʊj^{h}e
2Sg.Nom company.F Loc.in place.F give.Imper 1Sg.Acc

mera zɪmmɑh dek^h kɑr gɑr
1Sg.Gen.M responsibility see Conj if

koi bɑtla de mʊj^{h}e
any tell go.Subj 1Sg.Acc

'I am so feeble that if you let me sit in your company; it is on me if someone could recognize me by looking around.' (Asadullah Ghalib)

Sometimes the complementizer form is coreferenced with the demonstrative phrase. In (28a) *ıs tarah* ‘this way’ is coreferenced with the complementizer clause, likewise demonstratives link with the complementizer clauses in (28b–c).

(28) a. گیا تو اس طرح گیا کہ مدتوں نہیں ملا

gɑya	to	**ıs**	**tarɑh**	gɑya
go-Perf.M.3Sg	then	this	way	go-Perf.M.3Sg

kıh	muddɑtõ	nɑhĩ	mıl-a
that	period.F.3Pl	not	meet-Perf.M.3Sg

‘He went in such a way that he has not met since long.’ (Munir Niazi)

b. علی کا ندا کی ہر بات ماننا اس بات کا ثبوت ہے کہ وہ اس سے محبّت کرتا ہے.

ɑli=ka	nıda=ki	hɑr	bat	manna
Ali.M.3Sg=Gen	Nida.F.3Sg=Gen	every	demand	accept-Inf

ıs	**bat**=ka	sɑbut	hɛ	**kıh**
this.Sg	thing.F=Gen	proof.M.3Sg	be.Pres.M	that

vʊh	ʊs=se	mʊhɑbbɑt	kɑr-ta	hɛ
3Sg.Nom	3Sg=Abl	love.F	do-Imperf.M.3Sg	be.Pres

‘Ali’s accession to each of Nida’s demand is the proof that he loves her.’

c. علی کی ندا سے محبّت کا ثبوت یہ ہے کہ وہ اس کی ہر بات مانتا ہے.

ɑli=ki	nıda=se	mʊhabbɑt=ka	sɑbut	**yıh**	hɛ	**kıh**
Ali.M=Gen	Nida.F=Abl	love.F=Gen	proof.M	this	be.Pres	that

vʊh	ʊs=ki	hɑr	bat	man-ta	hɛ
3Sg.M	3Sg.F=Gen	every	demand.F	accede-Imperf.M.3Sg	be.Pres

‘Proof of Ali’s love with Nida is that he accedes to each of her demands.’

Due to these so many uses of the same form *kıh*, it is very hard to automatically identify the canonical complementizer clause for verbs.

2.2.5 Argument attachment ambiguities

Syntactically not only do verbs subcategorize for arguments in Urdu, but nouns and adjectives do too, as in many other languages. These nouns and adjectives in Urdu

usually are derived from verbal stems. So, an identified case phrase in a sentence might not be the argument of the verb. To automatically determine whether some case phrases actually are part of the noun or the adjective modifying the noun or the verb in a sentence of an unannotated corpus is not an easy task without incorporating a grammar.

(29) a. ندا نے زکام سے بچاؤ کی دوائی خریدی.

nɪda=ne [[zʊkam=se bɑcao]=ki dɑvai] xɑrid-i
Nida=Erg [[flu=Abl protection=Gen.F medicine.F] buy-Perf.F
'Nida purchased medicine for protection from flu.'

b. ندا نے بازار سے زکام کی دوائی خریدی.

nɪda=ne bazar=se [zʊkam=ki dɑvai] xɑrid-i
Nida=Erg bazar=Abl [flu=Gen.M medicine.F.Sg] buy-Perf.F
'Nida purchased medicine for flu from the market.'

In (29a), *davai* 'medicine' is the nominative argument of the verb *xarid* 'to buy'. This noun has a genitive modifier *bacao* 'protection' that itself takes a *se*-marked argument. So, the case marked arguments found in a sentence are not always the arguments of the main verb in the sentence. In (29b), however, the *se*-marked noun is not part of any noun. Here it should be considered as an adjunct of the verb *xarid* 'to buy'. Consider another example in (30) where a verb *mɪl* 'get' and a noun *cʰuṭ* 'exemption' are used, both of which can take an ablative marked argument.

(30) a. ندا کو امّی سے کافی رقم ملی.

nɪda=ko ɑmmi=se kafi rɑqɑm mɪl-i
Nida.F.Sg=Dat mother.F.Sg=Abl enough money.F.Sg get-Perf.F.Sg
'Nida got enough money from the mother.'

b. ندا نے استانی کو زبانی امتحان سے چھوٹ کے لئے کہا.

nɪda=ne	ʊstani=ko	zɑbani	ɪmtɪhan=se
Nida.F.Sg=Erg	teacher.F.Sg=Acc	verbal	examination.M.Sg=Abl

cʰuṭ=ke	lɪye	kɑh-a
exemption.F.Sg=Gen	for	ask-Perf.M.Sg

'Nida asked the teacher for exemption from the verbal examination.'

c. ندا کو زبانی امتحان سے چھوٹ مل گئی.

nɪda=ko	zɑbani	ɪmtɪhan=se	cʰuṭ
Nida.F.Sg=Dat	verbal	examination.M.Sg=Abl	exemption.F.Sg

mɪl	gɑ-i
get	go-Perf.F.Sg

'Nida got exemption from the verbal examination.'

d. صدر کو مقدمات سے استثنی مل گیا.

sɑdɑr=ko	mʊqɑddɑmat=se	ɪstɪsna
president.3Sg=Dat	court-case.3Pl=Abl	exemption.M.Sg

mɪl	gɑ-ya
get	go-Perf.M.Sg

'The president got [exemption from court-cases].'
'The president got [exemption] through court-cases.'

In (30a) the ablative marked argument is no doubt licensed by the verbal semantics attached with the verb and in (30b) it is attached with the noun *cʰuṭ* 'exemption'. In (30c), syntactically the attachment of an ablative marked argument is ambiguous. It can be attached to either the verb *mɪl* 'get' or the noun *cʰuṭ* 'exemption'. A more illustrative example with two readings is given in (30d). Other different types of attachment ambiguities in Urdu have been discussed in somewhat detail in Rizvi (2007).

2.3 Summary

In this chapter, different challenges faced by the acquisition of subcategorization information for verbs from a raw Urdu corpus have been explored. Due to a rich variety of

verbs in Urdu, it is not straightforward to identify which combination in the sentence actually is acting as the main verb. The position of arguments in the sentences is not helpful as Urdu is a free word order language, although the verb usually appears last in the clause. The case clitics cannot be the direct cues for the arguments as a single clitic is used to mark many case functions and a grammatical function can bear different cases in different contexts and constructions. Argument attachment is also a challenge. With these problems and challenges at hand, an algorithm is developed to extract frames of verbs from a raw Urdu corpus. This is discussed in the next chapter.

Chapter 3

Automatic lexical acquisition

This chapter describes an approach to inferring syntactic frames for verbs in Urdu from an untagged raw corpus. As already shown in Chapter 2, there is not always a one to one correspondence between case clitic form and case, and case and grammatical function in Urdu. Case clitics, therefore, cannot serve as direct cues for extracting the syntactic frames of verbs. So, an indirect two-step approach has been devised and implemented. In a first step, all case clitic combinations for a verb are extracted and the unreliable ones are filtered out by applying inferential statistics. In a second step, the information of occurrences of case clitic forms in different combinations as a whole and on an individual level is processed to infer all possible syntactic frames of the verb.

The chapter is organized as follows: section 1 describes the selection of the corpus for automatic subcategorizarion acquisition, section 2 describes how to identify verbs in the corpus, section 3 introduces types of subcategorization frames (SCFs) to be extracted. The subcategorization acquisition system for Urdu (SASU) is detailed in section 4. Results and evaluations are provided in section 5. The usability of the SASU system is described in section 6 and its limitations are described in section 7. Section 8 concludes the chapter.

3.1 Corpus selection

3.1.1 Corpora used in previous works

Corpus selection is the first necessary step in the automatic acquisition of SCFs. Acquisition results are influenced by many attributes of the selected corpus, like size, genre, annotation level, age of the corpus, etc. Size and annotation level of the corpus, however, matter much than other attributes.

Roland & Jurafsky (2002) have investigated the influence of corpus choice on acquisition results. They acquired subcategorization frames from five different corpora; two were developed from psychological experiments in which participants were asked to produce sentences, two corpora were subparts of the Brown corpus and the Wall Street Journal corpus (Marcus *et al.*, 1993), and one corpus developed from telephone conversations. They reported differences among frequencies of different frame types. They noted two major sources of differences: (a) the discourse type, and (b) the semantic choices, i.e. the word senses presented in the corpora.

Better results are expected with a corpus of a sufficiently huge size comprising all genres. Hence, the computational utilities are usually run on the maximum available data. In the last two decades corpora of different sizes have been used for SCF acquisition. At the early stages of lexical acquisition for English language, Brent (1993) and Ushioda *et al.* (1993) both used the corpus of Wall Street Journal (WSJ) but of different sizes i.e. 2.6 million words and 600,000 words respectively. Manning (1993) used the corpus of the New York Times newswire of 4 million words and Briscoe & Carrol (1993) used 1.2 million words of the Susanne corpus, the corpus of spoken English (SEC), and the Lancaster-Oslo-Bergen Corpus (LOB). Researchers later on used corpora of much bigger sizes. For example, Carrol *et al.* (1998) used the whole British National Corpus (BNC) of 117 million words.

The corpora used for SCF acquisition also differ in the level of annotation. Brent (1993) used a raw English corpus for subcategorization acquisition. Annotated part of speech (POS) tagged corpora were used by Manning (1993) and Ushioda *et al.* (1993) and, later on, English corpora parsed partially or fully (Briscoe & Carrol 1997; Kinyon & Prolo 2002; O'Donovan *et al.* 2005) were used.

3.1.2 Urdu corpus

Due to not having refined corpora of Urdu at hand, I have developed and used a raw Urdu corpus for extracting subcategorization information. This corpus was collected from Urdu websites, mainly from the Urdu newspaper *Roznama Jang* website and the BBC Urdu website. In a first step the corpus was cleaned to remove html tags and other spurious text not in the form of sentences.

There are some Urdu characters that were composed by two symbols in Unicode before. But now all single forms of Urdu alphabets have been assigned their own codes. These characters are shown in Table 3.1. Such differences of single form and composed form found in the corpus were normalized.

There are some Arabic-script characters that have almost same form in Arabic and Urdu but different encodings in Arabic and Urdu. These characters are listed in

Nr.	Letter	Single Form	Composite Form
i.	Alif with madda above	آ [D8A2]	آ [D8A7 + D993]
ii.	Hey with hamza above	ۂ [DB82]	ۂ [DB81 + D9B4]
iii.	Vao with hamza above	ؤ [D8A4]	ؤ [D988 + D9B4]
iv.	Choti ye with hamza above	ئ [DB80]	یٔ [DB8C + D9B4]
v.	Hamza connected bari ye	ۓ [DB93]	ئے [D8A6 + DB92]

Table 3.1: Single and composite forms of characters

Table 3.2. Some Urdu characters with different encodings were found in the corpus and were changed to their corresponding Urdu encoding forms.

Nr.	Letter	Arabic Form	Urdu Form
i.	Tey marbuta	ة [D8A9]	ۃ [DB83]
ii.	Kaf	ك [D983]	ک [DAA9]
iii.	Hey	ه [D987]	ہ [DB81]
iv.	Hey with hamza above	ۀ [DB80]	ۂ [DB82]
v.	Yey	ي [D98A]	ی [DB8C]
vi.	Alif maksura	ى [D989]	ی [DB8C]

Table 3.2: Multi-encoding characters

In addition to this, different spellings of some key words used in the corpus were also unified. The corpus had many segmentation errors as well. The words at the end of most sentences where a verb is found were not properly separated. To make it possible to have a maximum possible candidate sentences for the target verb, the corpus was properly segmented automatically. In the segmenting module, an Urdu lexicon of more than 60,000 words developed by CRULP was used to identify correct words.

The corpus that is used in our system after preprocessing has about 10 million words (276825 sentences). The size of this corpus, however, is not comparable to that of corpora that have recently been used in the subcategorization acquisition systems.

3.2 Identification of verbs

3.2.1 Different methods of identifying verbs

Subcategorization acquisition systems also differ in the identification of verbs. Brent (1993) recognized verbs in the corpus by finding those tokens that come with and without the suffix '-ing'. Tokens found in some specific lexical context, e.g. which directly follow a determiner, were filtered out. Ushioda *et al.* (1993) identified verbs from the part-of-speech tag and Manning (1993) recognized tokens used with auxiliaries by a finite state parser. In other approaches verbs were identified from the parses of sentences.

In our case, verbs are identified by comparing different conjugation forms of a verb with different tokens of sentences of the corpus. Most of the verbal stems in Urdu are conjugated in a regular pattern for tense, aspect, number and gender.

3.2.2 Urdu Verb Conjugator

Any regular stem can have upto 16 conjugation forms, among them three are for infinitives. For example, the conjugation forms of the four stems i.e. *dık^h* 'appear', *dek^h* 'see', *dık^ha* 'show' and *dık^hva* 'make some one show' derived from the root *dık^h* in Urdu are listed alongside their glossed interpretation in Table 3.3. The table shows that all four stems are inflected similarly. The two versions of causative stems were already explained in section 2.1.1.2.

(1) a. آج چاند نہیں دکھا.
aj cãd nɑhĩ dıkʰ-a
today moon.M not appear-Perf.M.Sg
'The moon did not appear today.'

b. ندا کو جنگل میں ایک سانپ دکھا.
nɪda=ko jãgɑl=mẽ ek sãp dɪk^h-a
Nida=Dat jungle.M=Loc.in one snake.M.3Sg appear-Perf.M.Sg
'Nida saw a snake in a jungle.'

The base form stem *dɪk^h* 'appear' in Table 3.3 is listed as an intransitive stem (1a) which although with a dative subject can be considered as a two-place predicate (1b).

Nr.	Intrans.	Trans.	Caus. I	Caus. II	Gloss
i.	dɪk^h-na	dek^h-na	dɪk^h-a-na	dɪk^h-va-na	Infin.M.3Sg.Dir
ii.	dɪk^h-ne	dek^h-ne	dɪk^h-a-ne	dɪk^h-va-ne	Infin.M.3Sg.Obl Infin.M.3Pl
iii.	dɪk^h-ni	dek^h-ni	dɪk^h-a-ni	dɪk^h-va-ni	Infin.F
iv.	dɪk^h-ta	dek^h-ta	dɪk^h-a-ta	dɪk^h-va-ta	Imperf.M.Sg
v.	dɪk^h-te	dek^h-te	dɪk^h-a-te	dɪk^h-va-te	Imperf.M.Pl
vi.	dɪk^h-ti	dek^h-ti	dɪk^h-a-ti	dɪk^h-va-ti	Imperf.F.Sg
vii.	dɪk^h-tĩ	dek^h-tĩ	dɪk^h-a-tĩ	dɪk^h-va-tĩ	Imperf.F.Pl
viii.	dɪk^h-a	dek^h-a	dɪk^h-a-ya	dɪk^h-va-ya	Perf.M.Sg
ix.	dɪk^h-e	dek^h-e	dɪk^h-a-e	dɪk^h-va-e	Perf.M.PL Subjn.3Sg
x.	dɪk^h-i	dek^h-i	dɪk^h-a-i	dɪk^h-va-i	Perf.F.Sg
xi.	dɪk^h-ĩ	dek^h-ĩ	dɪk^h-a-ĩ	dɪk^h-va-ĩ	Perf.F.PL
xii.	dɪk^h-ẽ	dek^h-ẽ	dɪk^h-a-ẽ	dɪk^h-va-ẽ	Subjn.Pl
xiii.	dɪk^h-ye	dek^h-ɪye	dɪk^h-a-ɪye	dɪk^h-va-ɪye	Subjn.Pl
xiv.	dɪk^h-yo	dek^h-ɪyo	dɪk^h-a-ɪyo	dik^h-va-ɪyo	Subjn.Sg.Fam
xv.	dɪk^h-o	dek^h-o	dɪk^h-a-o	dɪk^h-va-o	Subjn.Sg.Fam
xvi.	dɪk^h	dek^h	dɪk^h-a	dɪk^h-va	Subjn.Sg.Frk
xvii.	dɪk^h-ũ	dek^h-ũ	dɪk^h-a-ũ	dik^h-va-ũ	Subjn.1Sg

Table 3.3: Conjugations of different stems derived from the root *dɪk^h*

There are about 700 verbal roots in Urdu. Different verbal stems are derived form verbal roots and they are about 1200 in number. As a supplementary part of the SASU system, a verb conjugator module was developed to conjugate different stems. There are basically three paradigms of conjugations of a regular stem.

Paradigm 1: When a stem does not end in a vowel then just the affixes *ø, -na, -ne, -ni, -ta, -te -ti, -tĩ, -a, -e -i, -ĩ, ẽ, -ye, -yo, -o, ũ* are added to the stem to generate all forms of conjugations of the verb as is shown in Table 3.3.[1]

[1] The stem form itself acts as one form of conjugations. As in Urdu *ĩ* and *ẽ* are form identical due to the same middle shape of two *yes*, practically in Urdu script there are 16 conjugation forms of a regular stem; total 17 are listed in Table 3.3, (xi) and (xii) are form identical in Urdu script.

Paradigm 2: When a stem ends in the vowels *-i or -e*, the first eleven conjugation forms are generated by adding affixes *ø, -na, -ne, -ni, -ta, -te -ti, -tĩ, -a, -e, -˜* to the stem. For the 12th conjugation form, *hamza* and the affix *ẽ* are added to the stem. Other conjugations are generated by adding affixes *-jye, -jyo, -o, ũ* to the stem form.[2]

Paradigm 3: If a stem ends in the vowels *-a, -o, -u*, the affixes *ø, -na, -ne, -ni, -ta, -te -ti, -tĩ* are added to the stem to generate first eight forms of conjugations, and the 9th form is obtained by adding the affix *-a* and inserting the consonant *-y* in between.[3] The rest of conjugation forms are generated by adding affixes *-e -i, -ĩ, ẽ, -ye, -yo, -o, ũ* to the stem and inserting *hamza* in between.[4]

There are a very few verbs in Urdu whose conjugation forms, especially the perfect participles, are generated irregularly. Examples are *kar* 'do' and *ja* 'go'.

The verb conjugator module has been developed to conjugate all regular and irregular stems. The conjugator has been implemented by encoding the rules for inflections of the verb stems of different inflectional paradigms as described above. To find candidate sentences of a target verb in the corpora, all conjugations of the verb are generated by the conjugator module and then these conjugation forms are compared with tokens of sentences in the corpus.

3.3 Types of SCFs

Number and types of subcategorization frames are not always the same in different acquisition systems. Some systems extract frames at a higher level of abstraction, and others on a fine-grained level, e.g. making distinction between different types of clauses and different types of adpositional phrases. A distinction between arguments and adjuncts is also not considered in all systems. Some systems choose many subcategorization frames comprising both arguments and adjuncts while others choose only a subset of relevant frames which comprise only arguments for investigation. The flexible and non-restricted choice of subcategorization frames can be more useful in the sense to induce domain-independent and unforseen inforamtion. Other approaches are good for theory-related and domain-specific subcategorization knowledge. The choice depends on the size and genre of the corpus, on the NLP task which will use the extracted lexicon and of course on individual decisions as well.

[2] In Urdu script the vowel *-i* and the consonant *-y* form are identical, therefore the same form is pronounced either as consonant or as vowel.

[3] In some verbs, however, just the affix *-a* is added as in *hu-a* from *ho* 'be' and $c^{h}u$*-a* from $c^{h}u$ 'touch'.

[4] As in Urdu orthography two vowels cannot come together, therefore either the consonant *-y* or *hamza* are inserted between two vowels.

Existing approaches differ much in this respect. Brent (1993) acquired six types of frames for verbs in which direct objects, subcategorized clauses and infinitives were accomodated. Ushioda *et al.* (1993) also investigated six frame types considering a larger variety of complement types and making no distinction between adjuncts and arguments. Manning (1993) extracted information for 19 frame types without distinguishing between arguments and adjuncts. Briscoe & Carrol (1997) investigated 163 frame types with a fine-grained distinction of prepositional phrase types and arguments and adjuncts. Carrol *et al.* (1998) devised a grammar to extract all verb-adjacent constituents as frame types.

Arguments and adjuncts of a verb are treated differently in formal theories of syntax. Therefore, making a distinction between them is important while building a sophisticated lexicon of verbs. The following subsection gives a brief description of distinction between arguments and adjuncts.

3.3.1 Distinction between arguments and adjuncts

Arguments and adjuncts have been distinguished by many linguists (see, for example, Grimshaw 1990; Meyers *et al.* 1996; Pollard & Sag 1987). The following points made by Grimshaw (1990) best describe the difference between arguments and adjuncts:

i. Arguments can be selected and subcategorized, in the sense that their presence and the form they take are under the control of individual predicates.

ii. Arguments must be licensed: they can occur only if they are theta-marked by a predicate as a function of the predicate's argument structure.

iii. Adjuncts are not theta-marked and do not need to be licensed by relationship to an a-structure. They are not subcategorized. Hence, their form is free, and they are never required by a-structure.

A theta-marked argument is that participant of a clause which has some thematic role that is it is an agent or patient, etc. The set of theta-marked participants of a verb is called its argument structure (a-structure).

In a lexical entry, if some verb's adjunct is defined as an argument of the verb then the parser would generate a spurious parse or would not generate the parse at all of the valid input sentence. And if an argument of some verb is considered an adjunct and not listed in the lexical entry of the verb, then the parser may generate a spurious parse. Due to wrong lexical entries respecting subcategorization frames, the conditions of completeness and coherence could not be met in LFG framework and a valid sentence may not be parsed.

According to Meyers *et al.* (1996), argument/adjunct distinction for a parser of a natural language is important due to at least following four factors:

I Complements of a verb V occur with V more frequently than with other verbs, whereas adjuncts occur with equal frequency with a large variety of verbs;

II Incorrectly classifying a complement as an adjunct may cause a parser to miss a parse;

III Incorrectly classifying an adjunct as a complement may cause a parser to add a spurious parse;

IV In an accurate representation of predicate argument structure, heads predicate of their complements, but adjuncts predicate of the heads they modify.

Sometimes, it becomes difficult to distinguish optional arguments from adjuncts. Meyers *et al.* (1996) reported different criteria for an XP in a verbal phrase to be a complement (argument) or an adjunct. They defined complements and adjuncts of a verb as follows:

Complement: Given a Verb Phrase which includes a head verb V, a phrase XP, XP is a complement if XP is an intrinsic part of the action, state, event, etc. described by the VP, i.e., the verb predicates of the XP.

Adjunct: Given a Verb Phrase VP which includes a head verb V, a phrase XP, XP is an adjunct if XP modifies V or VP. XP is not intrinsic to VP.

Based on above definitions, linguistic theory and empirical observations Meyers *et al.* devised sufficient criteria and some rules of thumb for recognizing arguments and adjuncts of a verb. These hints for arguments and adjuncts are listed in Tables 3.4 and 3.5 respectively. Many points of their criteria are valid cross linguistically.

With this distinction of arguments and adjuncts in general, now I move on to discuss different types of arguments and adjuncts in Urdu

3.3.2 Arguments and adjuncts in Urdu

3.3.2.1 Case marked NPs as arguments/adjuncts

Some typical examples of case marked arguments and adjuncts in Urdu are given in (2)–(6).

Sufficient criteria for a complement	
1. Obligatoriness	XP is obligatory for VP to be grammatical or for a particular sense of V to be possible.
2. Passive	XP can only be the subject of the passive if XP is a complement. Only complement PPs can be stranded by pseudo passive.
3. Theta roles	XP has an argument theta role: theme, source, goal, patient, recipient, experiencer, proposition, question, etc.
4. Implied meaning	XP is optional, but is implied if omitted.
5. Selection restrictions	If V imposes selection restrictions on XP, XP is a complement.
Rules of thumb for a complement	
6. Frequency	XP occurs with verb V with high relative frequency.
7. Typical complements	NPs, PPs headed by "to", clauses (other than relative, "whether" and "if")
8. Complement alternation	XPs which participate in alternations are usually complements. (Levin, 1993)
9. Linear order	An XP between a head and a complement is probably a complement.
10. Island constraints	Most Complements can violate "island constraints"

Table 3.4: Hints for complement-hood

Criteria for an adjunct	
1. Frequency	XP occurs with most verbs with roughly the same frequency and meaning
2. Typical adjuncts	Purpose clauses, PPs/AdvPs/Subordinate clauses headed by "before", "after", "while", "because", "although", "if" or "by"; instrumental/concomitant "with" phrases, "by means of", benefactive, place, manner and time AdvPs and PPs .
3. Selection restrictions	An adjunct imposes selection restrictions on the verb/VP
4. WH Words	AdvPs/PPs which can be questioned with "Why" or "How"
5. Fronting adjunct	PPs front more naturally than complement PPs
6. Island constraints	Adjuncts cannot usually violate "island constraints"

Table 3.5: Hints for adjunct-hood

(2) a. ندا آج نہیں آئی.
nɪda aj nɑhĩ a-i
Nida.F.3Sg today not come-Perf.F.3Sg
'Nida did not come today.'

b. علی دوپہر کو آیا.
ɑli **dopɑhr=ko** a-ya
Ali.M.3Sg noon.Sg=Temp come-Perf.M.3Sg
'Ali came at noon.'

c. ندا گھر کو گئی.
nɪda **gʰɑr=ko** gɑ-i
Nida.F.3Sg home.M.Sg=Loc go-Perf.F.3Sg
'Nida went home.'

In (2a) the unmarked noun *nɪda* is the subject of the verb *a* 'come' and the unmarked noun *aj* is the adjunct in the clause. Temporal phrases can appear with almost all verbs; hence are treated as adjuncts. The *ko* marked participant in (2b) shows the time of the event and is an adjunct. In (2c), however, the *ko* marked participant is the destination/goal (thematic argument) of the verb *ja* 'go' and hence is considered as the complement of the verb rather than an adjunct.

(3) a. علی نے لائبریری میں کتاب پڑھی.
ɑli=ne **laɪbreri=mẽ** kɪtab pɑɽʰ-i
Ali.M.3Sg=Erg library.F.3Sg=Loc.in book.F.3Sg read-Perf.F.3Sg
'Ali read a book in the library.'

b. ندا نے غبارے میں ہوا بھری.
nɪda=ne **ɣʊbare=mẽ** hɑva bʰɑr-i
Nida.F.3Sg=Erg balloon.M.3Sg=Loc.in air.F.3Sg fill-Perf.F
'Nida filled air in the balloon.'

The *mẽ* marked nouns in both instances of (3) are locations. In (3a) the location is an adjunct of the verb *pɑɽʰ* 'read' and the sentence is grammatically valid without it. In

(3b), however, the sentence will become ungrammatical without the phrase *ɣʊbare=mẽ* ‘in the balloon’. Hence, here it is considered as an argument of the verb *b^{h}ar* ‘fill’.

(4) a. علی صبح سے پڑھ رہا ہے.

ɑli	**sʊbʊh=se**	pɑṛ^{h}	rɑh-a	hɛ
Ali.M.3Sg	morning.F.3Sg=Abl	study	Prog-Perf.F.3Sg	be.3Sg.Pres

‘Ali has been studying since morning.’

b. ندا نے علی سے کتاب مانگی.

nida=ne	**ɑli=se**	kɪtab	mɑ̃g-i
Nida.F.3Sg=Erg	Ali.M.3Sg=Abl	book.F.3Sg	ask-Perf.F

‘Nida asked Ali for the book.’

Both examples in (4) contain *se* marked participants. The *se* marked noun in (4a) is temporal, hence is an adjunct. In (4b), the *se* marked noun is not optional and should be treated as an argument of the verb. Removing this participant would make the sentence of the verb *mɑ̃g* ‘ask for’ ungrammatical.

(5) a. علی نے کارڈ پر ندا کا نام لکھا.

ɑli=ne	**kaṛḍ=pɑr**	nɪda=ka	nam	lɪk^{h}-a
Ali.M=Erg	card.M.Sg=Loc.on	Nida.F=Gen	name.M.Sg	write-Perf.F.3Sg

‘Ali wrote Nida’s name on the card.’

b. ندا نے کارڈ میز پر رکھا.

nɪda=ne	karḍ	**mez=pɑr**	rɑk^{h}-a
Nida.F.3Sg=Erg	card.M.3Sg	table.F.3Sg=Loc.on	put-Perf.M.3Sg

‘Nida put the card on the table.’

Locative marked nouns in Urdu are adjuncts for many verbs as in (5a). However some verbs require locative marked nouns obligatorily as in (5b). Without the case marked phrase *mez=par* ‘on the table’, the sentence will be not valid.

(6) علی اور ندا دریا کے پل پر جا رہے ہیں۔

ɑli	ɔr	nɪda	dɑrya=ke	**pʊl=pɑr**	ja
Ali.M.3Sg	and	Nida.F.3Sg	river.M.3Sg=Gen.Obl	bridge.F.3Sg=Loc.on	go
rɑh-e	hɛ̃				
Prog-Perf.M.Pl	be.Pres.Pl				

‘Ali and Nida are going over the river’s bridge.’
‘Ali and Nida are going to the river’s bridge.’

Sometimes, there is an ambiguous reading between an adjunct and an argument. For example, in (6), *pʊl* ‘bridge’ can be a location where Ali and Nida are going over or it can be a goal or destination which Ali and Nida are heading to. In the first case there would be an adjunct reading, but in the second case it would be a complement reading. So, two entires of *ja* ‘go’ should be added in the lexicon; one with and other without the *par* marked phrase as an argument otherwise the parser would generate only the one parse of the sentence with an adjunct reading as is the case with other intransitive verbs.

3.3.2.2 Adposition marked NPs as arguments/adjuncts

Adpositional phrases can also act either as adjuncts or arguments in a clause. Consider two examples in (7). In (7a) the adpositional phrase *ali=ki tarah* ‘like Ali’ is an adjunct of the clause and in (7b) the adpositional phrase *nɪda=ki tarah* ‘like Nida’ is an argument of the verb *lag* ‘seem’.

(7) a. علی کی طرح ندا بھی محنت کرتی ہے۔

ɑli=ki	tɑrɑh	nɪda	b^hi	mɪhnat	kɑr-ti	hɛ
Ali.M.3Sg=Gen	like	Nida.F.3Sg	also	work-hard	do-Imperf.F	be.3Sg.Pres

‘Nida works hard like Ali.’

b. یہ لڑکی ندا کی طرح لگتی ہے۔

yɪh	lɑṛki	nɪda=ki	tɑrɑh	lɑg-ti	hɛ
this	girl.F.3Sg	Nida.F.3Sg=Gen	like	seem-Imperf.F.3Sg	be.3Sg.Pres

‘This girl seems like Nida.’

(8) a. ندا کے لیے ریاضی مشکل مضمون ہے۔

nɪda=ke	lɪye	rɪyazi	mʊʃkɪl	mɑzmun	hɛ
Nida.F=Gen	for	mathematics	difficult	subject	be.3Sg.Pres

'Mathematics is a difficult subject for Nida.'

b. علی ایک یورپی ٹی وی چینل کے لیے کام کرتا ہے۔

ɑli	ek	yorpi	ṭi-vi	cɛnɑl=ke	lɪye	kam	kɑr-ta	hɛ
Ali.M	one	European	TV	channel=Gen	for	work	do-Imperf.M	be.3Sg.Pres

'Ali works for an European TV channel.'

The adpositional phrase *nɪda=ke lɪye* 'for Nida' in (7a) is an adjunct and the adpositional phrase *ek yorpi ṭi-vi cɛnɑl=ke lɪye* 'for an European TV channel' is an argument. Adpositional arguments/adjuncts in Urdu are discussed in detail in Chapter 6.

3.3.2.3 Infinitival arguments/adjuncts

Now consider some examples of clauses involving infinitives in (9).

(9) a. ندا کو کار چلانا پڑی۔

nɪda=ko	**kar**	**cɑla-na**	pɑṛ-i
Nida.F.3Sg=Dat	car.F.3Sg	drive-Inf.M.3Sg	fall-Perf.F.3Sg

'Nida had to drive a car.'

b. ندا نے کار چلانا چاہی۔

nɪda=ne	**kar**	**cɑla-na**	cah-i
Nida.F.3Sg=Erg	car.F.3Sg	drive-Inf.M.3Sg	want-Perf.F.3Sg

'Nida wanted to drive a car.'

c. ندا نے کار چلانا سیکھی۔

nɪda=ne	**kar**	**cɑla-na**	sik^h-i
Nida.F.3Sg=Erg	car.F.3Sg	drive-Inf.M.3Sg	learn-Perf.F.3Sg

'Nida learnt driving a car.'

d. ندا کو کار چلانا آتی ہے.

nɪda=ko	**kar**	**cɑla-na**	a-ti	hɛ
Nida.F.3Sg=Dat	car.F.3Sg	drive-Inf.M.3Sg	come-Imperf.F.3Sg	be.3Sg.Pres

'Nida knows driving a car.'

e. ندا نے کار چلانا شروع کی.

nɪda=ne	**kar**	**cɑla-na**	ʃʊru	k-i
Nida.F.3Sg=Erg	car.F.3Sg	drive-Inf.M.3Sg	start	do-Perf.F.3Sg

'Nida started driving a car.'

f. ندا کار چلانے لگی.

nɪda	**kar**	**cɑla-ne**	lɑg-i
Nida.F.3Sg	car.F.3Sg	drive-Inf.Obl	start-Perf.F.3Sg

'Nida started driving a car.'

g. علی نے ندا کو کار چلانے دی.

ɑli=ne	nɪda=ko	**kar**	**cɑla-ne**	d-i
Ali.M.3Sg=Erg	Nida.F.3Sg=Dat	car.F.3Sg	drive-Inf.Obl	give-Perf.F.3Sg

'Ali let Nida drive a car.'

All of the above instances are examples of long distance agreement (Bhatt, 2005; Davison, 1988; Mahajan, 1989). The matrix verb in them agrees with the object of the embedded infinitive. In some dialects the infinitive itself also agrees with its object.

Butt (1995) has analyzed verbal combinations in permissive constructions like in (9g) as complex predicates based on tests of agreement, anaphora and control. She argues that the verb *de* 'give' in such constructions agrees with the object of the embedded infinitive, hence the infinitive with the verb *de* 'give' acts as a complex predicate. This agreement test is also applicable to constructions in (9a–e). So, based on her analysis, verbal combinations in all examples of (9a–e) are instances of complex predicates in which the main verb is in infinitival form. In (9f) the unmarked higher argument is always the subject with which the aspectual verb *lag* 'start' agrees. So, constructions like in (9f) can be analyzed as of a simple predicate *lag* 'start' in which case it will be supposed to subcategorize for an infinitival argument.

In (9e) above, the verb *ʃʊru kar* ‘start’ is already a complex predicate and if the infinitive is also analyzed to form another complex predicate with it then the resulting complex predicate will be the combination of total three predicates. Consider the recursive embedding of infinitives in (10). In each instance of (10) the matrix verb embeds an infinitive which itself embeds another infinitive. The matrix verb agrees with the object of the inner most infinitive. Such constructions have not been investigated and analyzed yet and I am not sure whether these even can be analyzed as complex predicates.

(10) a. ندا کو کار چلانا سیکھنا پڑی.

nɪda=ko	**kar**	**cɑla-na**	**sik^h-na**	pɑɽ-i
Nida.F.3Sg=Dat	car.F.3Sg	drive-Inf.M.3Sg	learn-Inf.M.3Sg	fall-Perf.F.3Sg

‘Nida had to learn driving a car.’

b. ندا نے کار چلانا سیکھنا چاہی.

nɪda=ne	**kar**	**cɑla-na**	**sik^h-na**	cah-i
Nida.F.3Sg=Erg	car.F.3Sg	drive-Inf.M.3Sg	learn-Inf.M.3Sg	want-Perf.F.3Sg

‘Nida wanted to learn driving a car.’

c. علی نے ندا کو کار چلانا سیکھنے دی.

ɑli=ne	nɪda=ko	**kar**	**cɑla-na**	**sik^h-ne**	d-i
Ali=Erg	Nida=Dat	car.F.3Sg	drive-Inf.M.3Sg	learn-Inf.M.3Sg	give-Perf.F.3Sg

‘Ali let Nida learn driving a car.’

(11) a. ندا کو سکول جانا چاہیے.

nɪda=ko	sɑkul	ja-na	cahɪye
Nida.F=Erg	school.M	go-Inf.M.3Sg	must.3Sg

‘Nida must go to school.’

b. ندا کو سکول جانا ہے.

nɪda=ko	sɑkul	ja-na	hɛ
Nida.F=Dat	school.M	go-Inf.M.3Sg	be.3Sg.Pres

‘Nida has to go to school.’

c. ندا نے سکول جانا ہے.
nıda=ne sɑkul ja-na hɛ
Nida.F=Erg school.M go-Inf.M.3Sg be.3Sg.Pres
'Nida is to go to school.'

d. (ندا!) کل سکول جلدی آنا.
(nıda) kɑl sɑkul jɑldi a-na
(Nida.F) tomorrow school.M early come-Inf.M.3Sg
'(Nida!) come to school earlier tomorrow.'

The example sentence in (10a) is an instance of a modal construction where an infinitive is used. Some more examples of modal constructions where infinitives are used are given in (11).

The examples involving infinitives, I have put so far, can be analyzed as to forming constructions of complex predicates and modals in Urdu. Linguists may have different opinions on the argument/adjunct status of infinitival elements in these examples. Now I put some more examples where the argument/adjunct status of infinitival elements is clear. Consider first the purpose clauses in (12).

(12) a. ندا کتاب لوٹانے (کے لیے) لائبریری گئی.
nıda **kıtab** **lɔṭa-ne**=(ke lıye) laıbreri gɑ-i
Nida.F.3Sg book.F.3Sg return-Inf.Obl=Gen for library.F go-Perf.F.3Sg
'Nida went to the library to return the book.'

b. ندا نے انعام جیتنے کے لیے محنت کی.
nıda=ne **ın'am** **jit-ne**=ke lıye mıhnat k-i
Nida.F=Erg prize.M.3Sg win-Inf.Obl=Gen for hard-work.F do-Perf.F.3Sg
'Nida worked hard to win the prize.'

The purpose clause in Urdu can be constructed by marking an infinitive with the postposition *lıye* 'for' as in (12) and this clause always acts as an adjunct of the main clause.[5] When the matrix verb is a movement verb then the infinitive of the purpose clause can optionally be unmarked although in oblique form as in (12a).

[5] Examples of adjunct clauses (gerundal or infinitival) in English given by Meyers *et al.* (1996) are following: i. He opened the door [to let the cat out]. ii. She lay [smiling at me].

Infinitival clauses in Urdu sometimes encode the theme argument of main verbs. In (13) some examples of two transitive verbs *kah* 'say' and *puc*^{h} 'enquire' are given.

(13) a. علی نے ندا سے سکول جانے کے لیے کہا.
ɑli=ne nɪda=se sɑkul ja-ne=ke lɪye kɑh-a
Ali.M.3Sg=Erg Nida.F.3Sg=Com school go-Inf.Obl=Gen for say-Perf.M.3Sg
'Ali asked Nida to go to school.'

b. علی نے ندا سے سکول جانے کو کہا.
ɑli=ne nɪda=se sɑkul ja-ne=ko kɑh-a
Ali.M.3Sg=Erg Nida.F.3Sg=Com school go-Inf.Obl=Dat say-Perf.M.3Sg
'Ali asked Nida to go to school.'

c. علی نے ندا سے سکول جانے کا کہا.
ɑli=ne nɪda=se sɑkul ja-ne=ka kɑh-a
Ali.M.3Sg=Erg Nida.F.3Sg=Com school go-Inf.Obl=Gen say-Perf.M.3Sg
'Ali asked Nida to go to school.'

d. علی نے ندا سے سکول جانے کا پوچھا.
ɑli=ne nɪda=se sɑkul ja-ne=ka puc^{h}-a
Ali.M.3Sg=Erg Nida.F.3Sg=Com school go-Inf.Obl=Gen enquire-Perf.M.3Sg
'Ali enquired Nida whether she would go to school.'

(14) a. ندا نے کتاب پڑھنے کا سوچا.
nɪda=ne kɪtab pɑṛ^{h}-ne=ka so-ca
Nida.F.3Sg=Erg book.F.3Sg read-Inf.Obl=Gen think-Perf.M.3Sg
'Nida thought of reading the book.'

b. علی نے ندا کے سکول جانے کا معلوم کیا.
ɑli=ne nɪda=ke sɑkul ja-ne=ka mɑ'lum kɪ-ya
Ali.M=Erg Nida.F=Gen school go-Inf.Obl=Gen known do-Perf.M.3Sg
'Ali got to know whether Nida had gone to school.'

The infinitive participant of the verb *kah* ‘say’ is either marked by the postposition *lıye* (13a) or is marked for dative case (13b) or for genitive case (13c) and the infinitival participant of the verb *puc*ʰ ‘enquire’ is usually marked for genitive case (13d). In these verbs, the infinitival clauses, in fact, encode the theme of the verbs and hence are arguments. Genitive marked infinitival arguments of two more verbs are given in (14).

3.3.2.4 Clausal arguments/adjuncts

(15) a. ندا نے چاہا کہ وہ کار چلائے.

nıda=ne	cah-a	[kıh	vʊh	kar	cɑla-e]
Nida.F.3Sg=Erg	want-Perf.M.3Sg	that	3Sg.F	car.F.3Sg	drive-Subjn

‘Nida wanted to drive the car.’

b. ندا نے سوچا کہ کتاب پڑھی جائے.

nıda=ne	soc-a	kıh	[kıtab	pɑṛʰ-i	ja-e]
Nida.F.3Sg=Erg	think-Perf.M.3Sg	that	book.F.3Sg	read-Perf.F	go-Subjn

‘Nida thought of reading the book.’

c. علی نے ندا سے کہا کہ وہ سکول جائے.

ɑli=ne	nıda=se	kɑh-a	[kıh	vʊh	sɑkul	ja-e]
Ali.M.3Sg=Erg	Nida.F.3Sg=Com	say-Perf.M.3Sg	that	3Sg.F	school	go-Subjn

‘Ali asked Nida to go to school.’

In each each instance of (15), the *kıh* complementizer clause encodes the theme argument of the matrix verb. In (16) some examples are given which are equivalent to those of (12) in meanings. Both cases involve the purpose clauses. In (16), the *takıh* complementizer clauses encode the purpose and is an adjunct clause.

(16) a. ندا لائبریری گئی تاکہ کتاب لوٹائے.

nıda	laıbreri	gɑ-i	takıh	kıtab	lɔṭa-e
Nida.F.3Sg	library.F.3Sg	go-Perf.3Sg	so-that	book.F.3Sg	return-Subjn

‘Nida went to the library to return the book.’

b. ندا نے محنت کی تاکہ انعام جیتے.

nıda=ne mıhnɑt ki takıh ın'am jit-e
Nida.F.3Sg=Erg hard-work.F.3Sg do-Perf.3Sg so-that prize.M.3Sg win-Subjn
'Nida worked hard to win the prize.'

(17) a. ندا نے مسکراتے ہوئے چائے پی.

nıda=ne [mʊskɑra-te hu-e] cae pi
Nida.F=Erg smile-Imperf Conj.while tea.F drink-Perf.F.3Sg
'Nida took tea while smiling.'

b. ندا چائے پی کر سو گئی.

nıda [cae pi kɑr] so gɑ-i
Nida.F.3Sg tea.F.3Sg drink Conj.having sleep go-Perf.F.3Sg
'Nida slept after having taken tea.'

c. علی کے آتے ہی ندا جاگ گئی.

[ɑli=ke a-te hi] nıda jag gɑ-i
Ali.M=Gen come-Imperf Conj.when Nida.F awakw go-Perf.F.3Sg
'Nida awaked just when Ali came.'

There are many conjunctions in Urdu used to conjoin two clauses. The conjoined clauses always act as adjuncts, as shown in (17).

3.3.3 Number and types of SCFs considered for Urdu verbs

The number and types of subcategorization frames is drawn from the number and types of arguments in a language in general. Types of arguments of predicates can be defined on many bases. In Urdu, for example, arguments can be characterized in three ways: (a) based on types of case, the arguments are marked with, i.e. nominative, accusative, ergative, dative, comitative, etc., (b) based on grammatical functions of the participants, i.e. subject, direct object, indirect object, etc., and (c) based on thematic roles the participants are contributing to in the predicate, i.e. agent, patient, theme, source, goal, etc.

Consider some instantiations of the verb *lɪk^h* 'write' in Urdu in (18) with different numbers of arguments. Adjunct phrases are avoided in all these examples.

(18) a. علی نے خط لکھا.

ɑli=ne xɑt lɪk^h-a
Ali.M.3Sg=Erg letter.M.Sg write-Perf.F.3Sg
'Ali wrote a letter.'

b. علی نے ندا کو خط لکھا.

ɑli=ne nɪda=ko xɑt lɪk^h-a
Ali.M.3Sg=Erg Nida.F.3Sg=Dat letter.M.Sg write-Perf.F.3Sg
'Ali wrote a letter to Nida.'

c. علی نے اپنی بیماری بارے لکھا.

ɑli=ne ɑpni bimari bare lɪk^h-a
Ali.M.3Sg=Erg self illness.F.Sg about write-Perf.F.3Sg
'Ali wrote about his illness.'

d. علی نے ندا کو اپنی بیماری بارے لکھا.

ɑli=ne nɪda=ko ɑpni bimari bare lɪk^h-a
Ali.M.Sg=Erg Nida.F.Sg=Dat self illness.F.Sg about write-Perf.F.3Sg
'Ali wrote to Nida about his illness.'

e. علی نے لکھا کہ وہ بیمار ہے.

ɑli=ne lɪk^h-a kɪh vʊh bimar hɛ
Ali.M=Erg write-Perf.F.3Sg that.Compl 3Sg.Nom ill be.Pres.Sg
'Ali wrote that he was ill.'

f. علی نے ندا کو لکھا کہ وہ بیمار ہے.

ɑli=ne nɪda=ko lɪk^h-a kɪh
Ali.M=Erg Nida.F.Sg=Dat write-Perf.F.3Sg that.Compl
vʊh bimar hɛ
3Sg.Nom ill be.Pres.Sg
'Ali wrote to Nida that he was ill.'

The frames of the verb *likʰ* 'write' in (18) based on thematic roles of the participants are given as:

Nr.	Subcategorization Frame
(18a) i.	Agent Theme
(18b) ii.	Agent Goal Theme
(18c) iii.	Agent Theme
(18d) iv.	Agent Goal Theme
(18e) v.	Agent Theme
(18f) vi.	Agent Goal Theme

However, frameworks like LFG do not use the information of thematic roles to parse the sentence. Instead, they use information of grammatical functions of the participants required by a verb to parse the sentence. The grammatical functions in many languages are directly implied by the case type of the arguments.

The corresponding frames of the verb *likʰ* 'write' in (18) in terms of case/postposition type of arguments are given in the following:

Nr.	Subcategorization Frame
(18a) i.	Ergative/Nominative Nominative
(18b) ii.	Ergative/Nominative Dative Nominative
(18c) iii.	Ergative/Nominative Adpositional
(18d) iv.	Ergative/Nominative Dative Adpositional
(18e) v.	Ergative/Nominative Complementizer-clause
(18f) vi.	Ergative/Nominative Dative Complementizer-clause

As already explained in Chapter 2, different grammatical functions in Urdu can be marked for the same case and due to the free word order nature of the language the grammatical functions cannot be recognized by the position of the argument in contrast to other fixed order languages like English. So, the subcategorization information should be in terms of grammatical functions as follows.

Nr.	Subcategorization Frame
(18a) i.	Subject Direct-Object
(18b) ii.	Subject Indirect-Object Direct-Object
(18c) iii.	Subject Adpositional
(18d) iv.	Subject Indirect-Object Adpositional
(18e) v.	Subject Complementizer-clause
(18f) vi.	Subject Indirect-Object Complementizer-clause

Furthermore, one grammatical function of Urdu can be marked for different cases in different scenarios as it was shown in section 2.2.2. The verb lexicon of Urdu should

include both information of grammatical functions and case of arguments to provide rich information to the parser. So, I have considered arguments with information of both case and grammatical function. With this enriched information the frames of the verb are given in the following:

Nr.	**Subcategorization Frame**
(18a) i.	$\text{Subject}_{erg/nom}$ Object_{nom}
(18b) ii.	$\text{Subject}_{erg/nom}$ Object_{dat} Object_{nom}
(18c) iii.	$\text{Subject}_{erg/nom}$ $\text{Adpositional}_{about}$
(18d) iv.	$\text{Subject}_{erg/nom}$ Object_{dative} $\text{Adpositional}_{about}$
(18e) v.	$\text{Subject}_{erg/nom}$ Clause_{compl}
(18f) vi.	$\text{Subject}_{erg/nom}$ Object_{dative} Clause_{compl}

In the experiments, I have only considered case marked arguments and the complementizer clause for the subcategorization frames of a verb. The adpositional and infinitival arguments/adjuncts are not included in the SCF information. No restriction is made on different combinations of case marked arguments to be the valid frames for verbs. Our system itself would testify valid frames of the verbs.

3.4 Subcategorization Acquisition System for Urdu (SASU)

As already discussed, since we do not have more refined resources at hand for Urdu, I have developed a system (Raza, 2010a) that can infer subcategorization frames of verbs from an unannotated corpus. In the system, we test how many of the possible 64 case clitic form plus complementizer form combinations (CLCs) are valid for a verb. These different combinations are due to the presence or absence of five case clitics and one complementizer (2^6=64) in the candidate sentence of the target verb. Before testing a verb for these combinations, necessary screening is made for candidate sentences of the target verb. The screened sentences are delimited to the scope of the target verb and possible spurious case phrases are ignored. Using the hypothesis testing technique, different case clitic combinations are validated. By further processing the information of the valid case clitic combinations for a given verb, its subcategorization frames are inferred. This system differs from the previous work on other languages in that we recognize verbs in the corpus by matching corpus words to members of conjugation set of a given verb and induce frame types indirectly from frequencies of clitic combinations.

Many of the challenges faced with the automatic lexical acquisition from a raw Urdu corpus described in Chapter 2, have been addressed in the SASU system. This system differs from the existing approaches in frame acquisition method. I have first acquired valid case clitics and complementizer combinations and then frames are induced from

frequencies of these combinations applying some meta rules. Different components of the frame acquisition system for Urdu are shown in Figure 3.1 and are described in detail in the following subsections.

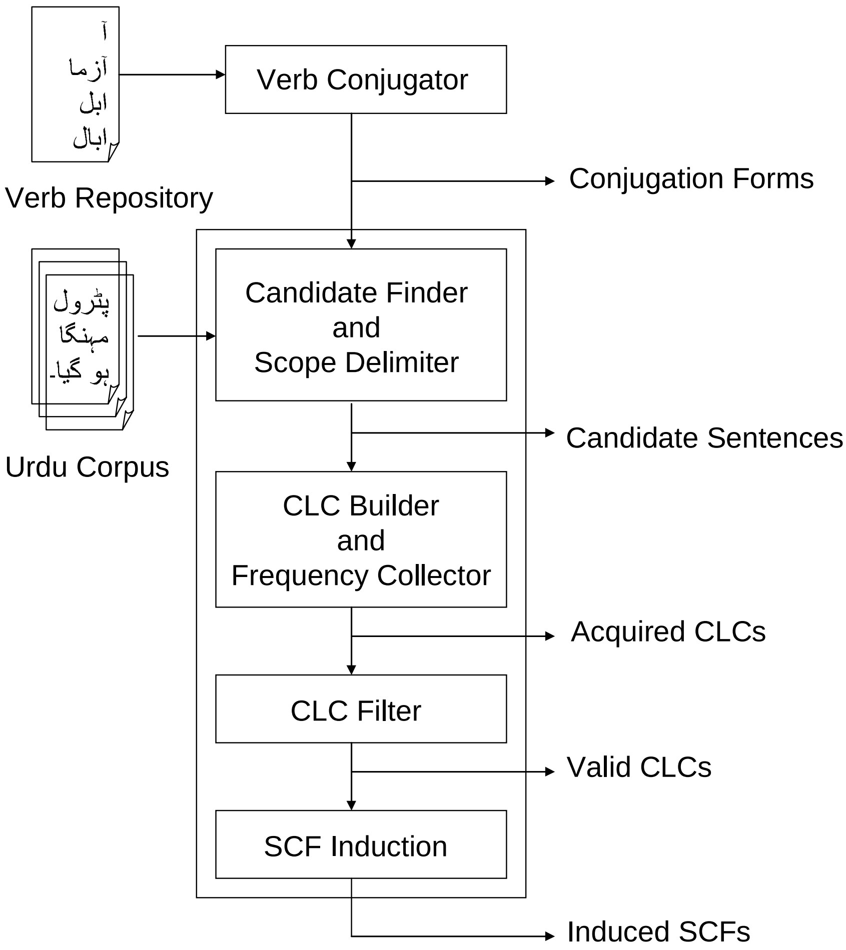


Figure 3.1: Block diagram of the SASU system

3.4.1 Candidate Finder and Scope Delimiter

The good candidate sentences for a target verb are extracted from the corpus in many steps. First, all the sentences where any of the conjugations of the target verb are found are extracted. Sometimes one conjugation of the target verb is either form-identical to some noun or some conjugation of another verb or some functional word. In such a

case those conjugations are not matched with the tokens of the sentences. After this initial step the screening and delimiting is made in three phases.

3.4.1.1 Initial screening phase

A simple verb in Urdu generally can act as a main verb or a light verb forming a complex predicate with some other predicate or even it can act as an aspectual auxiliary. The very first screening is made to make it sure that the verb has not been used in the sentence as a light verb or an aspectual auxiliary and that the verb is not found in a subordinating clause or passive construction. In this phase the candidate sentences are examined for the position and the context of the target verb. The corpus is passed through the following screens.

Screen 1: *Exclude those sentences where some other verb is found just before the target verb.*

In that case the target verb would have been used as a light or an auxiliary verb in the sentence rather as a main verb. Consider, for example, the verb *ḍal* 'put' in (19).

(19) a. علی نے سانپ مار ڈالا.

ɑli=ne sãp mar ḍal-a
Ali=Erg snake.M.Sg kill put-Perf.M
'Ali killed a snake.'

b. علی نے پانی میں کنکر ڈالا.

ɑli=ne pani=mẽ kɑnkɑr ḍal-a
Ali=Erg water=Loc pebble.M put-Perf.M
'Ali put a pebble in the water.'

In (19a) another verb *mar* 'kill' precedes the target verb *ḍal* 'put' and the target verb is functioning as a light verb of a complex predicate, rather than as a main verb. Here the obligatory locative marked argument of the verb *ḍal* 'put' does not come into play. In (19b) no other verb precedes the target verb and hence, the locative marked

argument of the target verb is there.

Screen 2: *The target verb should not be found in the subordinating clause of the sentence.*

This screen is inserted due to the fact that the arguments of the verb in subordinating clause could be controlled externally and the external argument may not match the canonical argument of the target verb in the subordinating clause. So, remove all the sentences where the target verb is followed by subordinating conjunctions *kar/ke*, *hi*, *hua/hue/hui*. Consider, for example, the case of (20) with the target verb *dɔṛ* 'run'.

(20) a. علی نے دوڑ کر پتنگ پکڑی.
ɑli=ne [dɔṛ kɑr] pɑtɑ̃g pɑkṛ-i
Ali=Erg run Conj.having kite.F.3Sg catch-Perf.F
'Ali caught a kite by running.'

b. علی دوڑا.
ɑli dɔṛ-a
Ali.Nom run-Perf.M
'Ali ran.'

The verb *dɔṛ* 'run' is an intransitive verb that does not take an ergative subject, but a nominative subject as in (20b). In (20a) the ergative subject is due to the main clause verb *pakaṛ* 'catch'. If a sentence like (20a) is not blocked to be a candidate sentence of the target verb *dɔṛ* 'run', then there is a chance that the system incorrectly infers an ergative subject for this verb.

Screen 3: *The target verb should not precede the* ja 'go' *auxiliary.*

The passive clause of a verb in Urdu is formed by the perfect participle of the verb followed by *jana* 'go'.[6] In a passive construction, the agent of the verb usually is

[6]The verb *cal* 'walk' is an exception to this rule.

demoted and is either left out or appears as a *se*-marked NP. Therefore such sentences should not be included in the set of candidate sentences of the target verb.

(21) a. کھانا کھایا گیا.

kʰana kʰa-ya gɑ-ya
meal.Nom eat-Perf.M go.Perf.M
'The meal was eaten.'

b. علی نے کھانا کھایا.

ɑli=ne kʰana kʰa-ya
Ali=Erg meal.Nom.M eat-Perf.M
'Ali ate the meal.'

In the passive construction (21a) the subject of the verb *kʰa* 'eat' is demoted and left out, while (21b) gives the information that this verb takes an ergative subject.

Some verbal stems form multiword verbs (complex predicates) with *ja* verb (light verb) and this phenomenon in Urdu affects the argument structure of the main verb. The verb *kʰa* 'to eat' that usually takes an ergative subject; in case of a complex predicate *kʰa ja* 'eat' it takes always a nominative subject as in (22).

(22) علی کھانا کھا گیا.

ɑli kʰana kʰa gɑ-ya
Ali.Nom meal.Nom eat.Perf.M go.Perf.M
'Ali ate the meal.'

As the argument structure of the main verb changes in both cases: when *ja* comes as a passive auxiliary or as a light verb after the main verb, therefore, we exclude all such sentences from the list of candidate sentences. There are other light verbs also that change the canonical argument structure of main verbs in Urdu. But in this work only the *ja* verb has been considered.

3.4.1.2 Scope delimiting phase

A sentence can have other clauses in it, for example: coordinating clause, relative clause etc. Each clause can have a different main verb. Delimitation of the candidate sentences to the scope of the target verb is made by identifying the following three patterns.

Pattern 1: Verb + *ɔr* || Verb + *ya* || Verb + *nah*

The words *ɔr* ‘and’, *ya* ‘or’ and *nah* ‘not’ are conjunctions that can conjoin two nouns or two clauses. If any of the above patterns is found in the forward direction after the target verb, then the sentence should be delimited to that point because in such a case another sentential clause might have been conjoined and share some arguments with the first clause. This is shown in different instances of (23) for three target verbs *xɑrid* ‘buy’, *a* ‘come’ and *ro* ‘weep’.

Arguments of the preceding clauses stay in their place when the the following clauses are removed. The advantage of this removal is that the system does not recognize the arguments of the second coordinating clause as arguments of the first clause. For example, the ergative and dative arguments of the second clause in (23b) would not be counted as the argument of the verb *a* ‘come’ in the first clause.

(23) a. علی نے احمد سے انگوٹھی خریدی اور ندا کو دی.

ɑli=ne	ɑhmɑd=se	ãguṭ^{h}i	xɑrid-i	**ɔr**	nɪda=ko
Ali.M=Erg	Ahmad.M=Abl	ring.F.3Sg	buy-Perf.F	and	Nida.F=Dat

d-i
give-Perf.F
‘Ali purchased a ring from Ahmad and gave it to Nida.’

b. علی اتفاقا آیا یا ندا نے اس کو بلایا تھا.

ɑli	ɪtfaqɑn	a-ya	**ya**	nɪda=ne	ʊs=ko	bʊla-ya	t^{h}a
Ali	by-chance	come-Perf.M	or	Nida=Erg	3Sg=Acc	invite-Perf.M	be.M.3Sg

‘Ali came by chance or Nida had invited him.’

c. ندا روئی نہ کسی کو مدد کے لیے بلایا۔
nɪda ro-i **nɑh** kɪsi=ko mɑdɑd=ke lɪye pʊkar-a
Nida.Nom weep-Perf.F not someone=Acc help=Gen for call-Perf.M.Sg
'Nida neither wept nor called someone for help.'

This pattern, however, is not tested in the backward direction i.e. before the target verb in this phase. If we discard the sentence up to that pattern before the target verb then some of arguments of the target verb clause could be deleted due to the deleted first conjunction clause. Consider again the case of the target verb *xarid* 'buy' in (24).

(24) a. علی نے بنک سے رقم نکلوائی اور گھڑی خریدی۔
ɑli=ne bɑ̃k=se rɑqɑm nɪklva-i **ɔr** g^hɑṛi xɑrid-i
Ali.M=Erg bank.F=Abl money.F draw-Perf.F and watch.F buy-Perf.F
'Ali draw cash from the bank and purchased a watch.'

b. علی نے گھڑی بیچی یا خریدی؟
ɑli=ne g^hɑṛi bec-i **ya** xɑrid-i
Ali.3Sg.M=Erg watch.F.Sg sell-Perf.F.3Sg or buy-Perf.F.Sg
'Did Ali buy or sell the watch?'

c. علی نے گھڑی بیچی نہ خریدی۔
ɑli=ne g^hɑṛi bec-i **nɑh** xɑrid-i
Ali.3Sg.M=Erg watch.F.3Sg sell-Perf.F.Sg or buy-Perf.F.Sg
'Ali neither sold nor purchased the watch.'

By removing the clauses before the conjunctions *ɔr* 'and', *ya* 'or' and *nah* 'not' in all three instances of (24) would also take out some of the arguments of the target verb *xarid* 'buy' and if the preceding clauses are not removed some extra or wrong arguments of the verb would be recognized. So, if such a pattern is found before the target verb, we exclude such a sentence as a whole in the final screening phase.

Pattern 2: Verb + any other conjunction or complementizer

The sentence is delimited to either side of the target verb where such a pattern is found. If the complementizer is found after the target verb then it is retained because it is a signal for the complementizer clause argument of the target verb. If it is found before the target verb, it is deleted as in that case it is not a signal that the target verb is taking complementizer clause argument. If such deletion is not made, the later module would wrongly consider it as the argument of the verb. Consider the use of a conjunction *jabkɪh* 'while' and the complementizer form *kɪh* 'that' in Urdu in (25).

(25) a. علی نے سنار کو اپنی انگوٹھی بیچی جبکہ ندا نے سنار سے ایک ہار خریدا.

ɑli=ne	sʊnar=ko	ɑpni	ɑ̃guṭ^{h}i	bec-i
Ali.M=Erg	jeweller.M=Abl	self	ring.F.3Sg	sell-Perf.3Sg.F

jɑbkɪh	nɪda=ne	sʊnar=se	ek	har	xɑrid-a
while	Ali.M.3Sg=Erg	jeweller.M=Abl	one	garland.F.3Sg	buy-Perf

'Ali sold his ring to the jeweller while Nida purchased a garland from him.'

b. علی نے کہا کہ اس نے ندا کو دیکھا ہے.

ɑli=ne	kɑh-a	**kɪh**	ʊs=ne	nɪda=ko	dek^{h}-a	hɛ
Ali.M=Erg	say-Perf.M	that	3Sg=Erg	Nida.F.Sg=Acc	see-Perf.M.Sg	be.Pres

'Ali said that he has seen Nida.'

c. علی نے دیکھا کہ ندا آ رہی ہے.

ɑli=ne	dek^{h}-a	**kɪh**	nɪda	a	rɑh-i	hɛ
Ali=Erg	see-Perf.M.3Sg	that	Nida	come	Prog-Perf.F.3Sg	be.Pres

'Ali saw that Nida was coming.'

The conjunction form *jabkɪh* 'while' is coordinating two complete clauses in (25a). Removing any of coordinating clause in such a case would not remove arguments of other clause. If we consider the verb *dek*^{h} 'see' as a target verb in (25b–c), the complementizer *kɪh* 'that' will not be included in the frame of the verb in (25b) as it actually is occurring in the outer clause which has introduced the clause in which *dek*^{h} 'see' is contained. But the complementizer form will be included in the frame of the verb *dek*^{h} 'see' in (25c) where it is found just after the target verb.

Pattern 3: Verb + any relative pronoun

For this pattern also, the sentence is delimited on both sides of the target verb. If the relative pronoun is found after the target verb, then it should be deleted as in such case it is the argument of some verb in the coming clause, else it is retained as it is in the domain of the target verb.

(26) a. اس سنار کو جس سے ندا نے انگوٹھی خریدی، علی نہیں جانتا.

ʊs	sʊnar=ko	[jɪs=se	nɪda=ne	ãguṭ^{h}i	xɑrid-i]
3Sg	jeweller=Acc	3Sg.RelP=Abl	Nida.F=Erg	ring.F	buy-Perf.F

ɑli	nɑhĩ	jan-ta
Ali.M.3Sg	not	know-Imperf.M.3Sg

'Ali does not know the jeweller who Nida purchased the ring from.'

b. ندا نے اس سنار سے انگوٹھی خریدی، جس کو علی نہیں جانتا.

nɪda=ne	ʊs	sʊnar=se	ãguṭ^{h}i	xɑrid-i
Nida.F=Erg	3Sg	jeweller.M.3Sg=Abl	ring.F.3Sg	buy-Perf.F.3Sg

[jɪs=ko	ɑli	nɑhĩ	jan-ta]
3Sg.Rel=Acc	Ali	not	know-Imperf.M.3Sg

'Nida purchased the ring from that jeweller whom Ali do not know.'

Two sentences of the target verb *xarid* 'buy' are given in (26) where in the first case the relative pronoun is found before the target verb (26a) and in the later case it is found after the target verb (26b). In the first case the tokens up to the relative pronoun would be truncated and the relative pronoun would be retained but in the second case the relative pronoun and the tokens after it both would be cut off. We note that in the first case extra arguments of the target verb could be detected due to presence of another verb. So, such type of sentences are dealt with in the final screening phase which is described in the next section.

3.4.1.3 Final screening phase

Although we have delimited the sentences to the scope of the target verb in the previous phase, still there might be other participle adjectives with their arguments within the scope of the target verb. So, to avoid counting a non-argument of the verb as an argument of the verb, we ignore such sentences.

Screen 4: *Ignore the sentences where any other verb before the target verb is found or after it after the light and/or auxiliary verbs.*

Consider the target verb *ʊṭʰa* ‘pick’ in (27).

(27) a. علی نے میز پر رکھے قلم کو اٹھایا.
ɑli=ne [mez=pɑr rɑkʰe] qɑlɑm=ko ʊṭʰa-ya
Ali=Erg table.F=Loc place.Perf.Obl pen.M=Acc pick-Perf.M
‘Ali picked the pen placed on the table.’

b. علی نے میز پر قلم کو رکھا اور اٹھایا.
ɑli=ne mez=pɑr qɑlɑm=ko rɑkʰ-a ɔr ʊṭʰa-ya
Ali=Erg table.F=Loc pen.M=Acc put-Perf.M and pick-Perf.M
‘Ali put the pen on the table and picked it’

c. علی نے قلم کو اٹھایا.
ɑli=ne qɑlɑm=ko ʊṭʰa-ya
Ali=Erg pen.M=Acc pick-Perf.M
‘Ali picked the pen.’

In (27a) *rakʰe* ‘placed’ is a participle adjective and takes a locative marked argument. In (27b) the verb *rakʰ* ‘place’ is a main verb and takes a locative marked argument. Such sentences (27a–b) are excluded so that the locative marked argument should not be inferred for the target verb *ʊṭʰa* ‘pick’, that does not take such argument as in (27c).

A limited number of adjectives are used in Urdu that are derived from Arabic verbal stems and do take case marked arguments. A few examples of such adjectives are *lahıq*

'attached', *mʊstasna* 'excepted', *ʃamɪl* 'included', *mʊʃtamɪl* 'comprised' taking *ko*, *se*, *mẽ* and *par* marked arguments respectively. Argument-taking adjectives are discussed in some detail in section 5.3. If such adjectives are found in candidate sentences, then the clitics associated with their arguments are ignored while acquiring CLCs of the target verb.

3.4.2 CLC Builder and Frequency Collector

Once the candidate sentences have been found, screened and delimited to the scope of the target verb, the counts for the different types of case clitics and complementizer combinations (CLCs) are computed. The type of a CLC is distinguished by the value of a six bit vector. These bits are for the five clitic forms *ne*, *ko*, *se*, *mẽ*, *par* and one complementizer form *kɪh* from the most significant bit to the least significant bit, respectively. For example, the vector value 110000 represents that CLC of the target verb for a sentence where only *ne* and *ko* clitics are found and others are absent. Depending upon the absence or presence of five clitics and one complementizer in the sentence, 64 CLC types are possible.

The counts of different CLC types for the verb *ʊṭʰa* 'pick' from its 248 final candidate sentences are given in Table 3.6. The CLCs of zero count are not mentioned.

3.4.3 CLC Filtering

The CLC types recognized in the last step may have some noise due to some wrong hits for the clitics. For filtering usually a null hypothesis (H_0) is formulated, which is assumed true unless there is some evidence to the contrary. An alternate hypothesis (H_1) is accepted in case some evidence proves H_0 false. Four methods reported in the literature can potentially be used for filtering out the unreliable CLCs.

3.4.3.1 Relative frequencies

One simple method is that relative frequencies of different CLC types be computed and if relative frequency of a CLC is higher than some threshold value then CLC is accepted else it is rejected. Lapata (1999) used this method to filter subcategorization frames for diathesis alternation detection. She used the COMLEX Syntax dictionary (Grishman *et al.*, 1994) to compute threshold frequency for each SCF from the the frequencies of SCFs in the dictionary. She reported that this method produced slightly better results than binomial filter discussed in section 3.4.3.4. Korhonen *et al.* (2000) also showed that this method performed better than binomial filter.

CLC Type (ne+ko+se+$m\tilde{e}$+par+$k\imath h$)	CLC Frequency
000000	72
100000	75
010000	08
001000	25
000100	19
000001	02
110000	04
101000	07
100100	14
100001	05
011000	01
010100	04
001100	04
111000	02
101100	02
100101	03
111100	01

Table 3.6: Types of CLCs and their counts recognized for the verb $\upsilon\d{t}^h a$ 'pick' with 248 candidate sentences

As we had no reference resource for setting a cut-off on the relative frequencies of CLCs, this method was not selected for filtering in our system.

3.4.3.2 Log likelihood ratio

The log likelihood ratio (LLR) reflects the difference between the observed and the expected distribution. As a null hypothesis, it could be assumed that the distribution of CLC is independent of the distribution of a verb that is $p(CLC|verb) = p(CLC)$. The LLR statistic verifies or rejects this hypothesis. The greater the LLR value, the more likely it is that the CLC is associated with the verb. If LLR is greater than some threshold value, then the null hypothesis is rejected. To calculate the LLR for each verb and CLC combinations four counts are required.

k_1 = Number of times CLC occurs with the verb
n_1 = Number of occurrences of the verb
k_2 = Number of times CLC occurs with any other verb
n_2 = Number of occurrences of other verbs

Using these numbers, the following three probabilities are computed:

$$p_1 = \frac{k_1}{n_1}, p_2 = \frac{k_2}{n_2}, p = \frac{k_1 + k_2}{n_1 + n_2}$$

Assuming that these probabilities are binomially distributed, the LLR statistic as given by Dunning (1993) is computed as:

$$\textit{-2log}\lambda = \textit{2[log L } (p_1,\ k_1,\ n_1) + \textit{log L } (p_2,\ k_2,\ n_2) \textit{ -}$$
$$\textit{log L } (p,\ k_1,\ n_1) \textit{ - log L } (p,\ k_2,\ n_2)]$$

where,

$$\textit{log L } (p,\ n,\ k) = k \textit{ log } p + (n\text{-}k)\textit{log}(1\text{-}p)$$

Sarkar & Zeman (2000) compared the LLR method with T-scores and binomial filter in their SCF acquisition system for Czech and showed that the F-measure was better with a binomial filter. And Korhonen *et al.* (2000) also showed that F-measure with binomial filter was better than the LLR method.

3.4.3.3 T-scores

T-scores can also be used to measure association between the CLC and the verb. Using the definitions from the previous section, T-score is computed as follows:

$$T = \frac{p_1 - p_2}{\sqrt{\sigma^2(n_1, p_1) + \sigma^2(n_2, p_2)}}$$

where,

$$\sigma(n, p) = np(1 - p)$$

The CLC will be assumed valid for the verb if T is greater than some threshold value.

3.4.3.4 Binomial filter

In a binomial filter Brent (1993) H_0 is formulated as the SCF (CLC in our case) is not associated with the verb and that the error probability (p_e) of occurring CLC with the verb is binomially distributed. Then probability of appearing m counts of CLCs in the total n occurrences of the verb can be computed by the following formula.

$$P(m, n, p_e) = \frac{n!}{m!(n-m)!} p_e^m (1 - p_e)^{n-m}$$

The following summation equation gives the probability of a CLC occurring m or more times.

$$P(m+, n, p_e) = \sum_{i=m}^{n} P(i, n, p_e)$$

If this probability $P(m+, n, p_e)$ is less than some threshold value, then H_0 is rejected and the CLC is assumed valid for the verb. A lesser threshold value gives a higher confidence level.

Briscoe & Carrol (1997) estimated error probabilities using reference resources. They defined verb classes based on frames and class membership probabilities of verbs were computed by dividing the number of verbs occurred with the class in Alvey NL Tools (ANLT) dictionary by total number of verbs in the dictionary. The probability of a pattern for class i was computed by dividing the number of patterns for class i extracted from the Susanne corpus by the total number of the patterns. The probability (p_e) of the verb not of class i occurring with a pattern of class i was computed by multiplying the complement of class member ship probability to the pattern probability as follows.

$$p_e = (1 - \frac{|verbs\ in\ class\ i|}{|verbs|})\frac{|patterns\ for\ i|}{|patterns|}$$

Brent, however, computed error probabilities experimentally from the corpus. In his method (see Brent (1993) for detail) a fixed number of first occurrences of verbs in the corpus are examined for different SCFs. For a specific SCF, first N occurrences of verbs each are distributed over histogram bins representing the number of times, a verb appeared with the SCF. A bin *i* contains the number of verbs appearing with the SCF exactly *i* times. The verbs that do not associate with the SCF are clustered towards the lower bins. Starting form the first bin to higher ones, the error probability is estimated that nearly fits to the binomial distribution.

In the experimentation, I used Brent's method to compute error probabilities by extracting CLCs of first 100 occurrences of 60 verbs in the corpus.

3.4.4 SCF Induction

Once the unreliable CLCs for a verb are filtered out, the SCFs for the verb are induced in the following three stages.

3.4.4.1 Application of Metarules

The information about the presence or absence of *ne* and *ko* bits in different CLCs is sufficient to infer information about the subject of the verb. The following three metarules are applied to infer what the subject of the verb is.

1. If the *ne* bit in a significant number of the CLCs is 1[7], then the subject of the verb is ergative/nominative and the verb is transitive.[8]

2. If both the *ne* and the *ko* bits are almost always zero then the subject of the verb is nominative and the verb is intransitive.[9]

3. If the *ne* bit is almost always zero and the *ko* bit is 1 in most of CLCs, then the subject of the verb is dative.

3.4.4.2 CLC Collapse

After the application of metarules we can collapse CLCs of the verb ignoring the bits in CLCs whose information has already been exploited. If the subject of the verb is inferred by applying Metarule 1 then ignoring the *ne* bit, 64 CLCs could be collapsed into 32 CLCs. Application of Metarule 2 leads to the collapse of 64 CLCs to 16, as here the two bits *ne* and *ko* are ignored. After applying Metarule 3 also, CLCs are collpased into 16. So, for three types of verbs, at the most 32, 16 and 16 number of SCFs respectively can be induced with the system.

CLC Type	**CLC Frequency**
000000+100000	72+75=147
010000+110000	08+04=12
000100+100100	19+14=33
001000+101000	25+07=32

Table 3.7: CLCs collapse for the verb *ʊṭʰa* ‘pick’

The collapse of the CLCs for the verb *ʊṭʰa* ‘pick’ is shown in Table 3.7.

3.4.4.3 SCF Information Collection

Different bits of CLCs after collapse give the exhaustive information about the SCFs of the verb. The number of these CLCs are actually the number of valid SCFs for the

[7]That is, counts of the CLCs where the *ne* bit appears and the counts of the CLCs where the *ne* bit does not appear are comparable.

[8]A very few intransitive verbs take a volition-based ergative subject. These are handled separately. It becomes more certain that an ergative subject is of a transitive verb if in some of the CLCs, a *ko* bit is also 1 alongside *ne* bit.

[9]A very few verbs are an exception to this rule. The transitive verb *la* ‘bring’ always takes the nominative subject. This is probably due to the reason that the verb *la* is the reduced form of the complex predicate *le a* ‘bring’ and the nominative subject is due to the intransitive light verb *a* ‘come’ in it.

verb. In case of the transitive verbs, the value 1 of the *ko* bit is the signal that verb takes accusative object and the 0 value of the *ko* bit tells that the verb takes nominative object. The other bits fill the rest of information about SCFs. For the intransitive verbs with nominative subject, other arguments in SCFs recognized are usually adjuncts of those verbs.

Nr.	**SCF**
1	Subject$_{erg/nom}$
2	Subject$_{erg/nom}$Object$_{acc}$
3	Subject$_{erg/nom}$Object$_{loc.in}$
4	Subject$_{erg/nom}$Object$_{abl}$

Table 3.8: SCFs induced for the verb *ʊṭʰa* 'pick'

The final subcategorization information induced for the target verb *ʊṭʰa* 'pick' is shown in Table 3.8. The ergative/nominative subject of the verb is induced by having significant number of CLCs with the *ne* bit as is shown in Table 3.7. The rest of arguments in the frames actually are the translation of the case clitic bits with value 1 in Table 3.7 to the grammatical functions. The case clitic *ko* bit is the signal of the accusative object and likewise the case clitic *se* bit is the signal of the ablative marked object.

3.5 Results and evaluation

Sixty basic verbs of Urdu were examined for CLCs and SCFs with the SASU system. In the first stage, we tested our CLC acquisition system. The reliable CLCs were filtered out by binomial hypothesis testing. The error probabilities were computed by the method as proposed by Brent (1993). The threshold value was set to 0.01 to achieve a 99% confidence level. The valid CLCs for different verbs were compared with hand judgments. The results for 22 CLCs have been displayed in Table 3.9. As was already mentioned that the *kɪh* form is multifunctional in Urdu, the CLCs results show that false positives are more when the *kɪh* bit is on in the CLCs.

In the second stage, the SCFs information was induced from the CLCs extracted in the first stage. In our system, SCF information includes both grammatical function and case of the argument of the verb. The subject has also been included in SCF information, as in Urdu subjects of different types of verbs are marked for different cases. Table 3.10 shows evaluation for 9 SCFs.[10]

[10]In Tables 3.9 and 3.10, TP=true positives; FP=false positives; TN=true negatives, FN=false

CLC	p_e(CLC)	TP	FP	TN	FN	MCV	Accuracy
000000	0.0000	59	1	0	0	1	98.34 %
100000	0.0023	38	6	13	3	9	85.00 %
010000	0.0137	32	2	17	9	11	81.67 %
001000	0.0093	36	5	13	6	11	81.67 %
000100	0.0025	44	11	3	2	13	75.00 %
000010	0.0000	0	0	54	6	6	90.00 %
000001	0.0077	13	19	28	0	19	68.34 %
110000	0.0021	19	4	17	20	24	60.00 %
101000	0.0044	12	3	39	6	9	85.00 %
100100	0.0009	37	7	16	0	7	88.34 %
100010	0.0000	0	0	57	3	3	95.00 %
100001	0.0021	8	8	39	5	13	78.34 %
011000	0.0050	10	2	47	1	3	95.00 %
010100	0.0024	19	7	34	0	7	88.34 %
010010	0.0000	0	0	56	4	4	93.34 %
010001	0.0018	11	9	38	2	11	81.67 %
001100	0.0058	15	7	36	2	9	85.00 %
001010	0.0000	0	0	58	2	2	96.67 %
001001	0.0059	5	3	50	2	5	91.67 %
000110	0.0000	0	0	60	0	0	100.00 %
000101	0.0022	11	6	43	0	6	90.00 %
000011	0.0000	0	0	57	3	3	95.00 %

Table 3.9: Results of 22 CLCs for 60 verbs compared with hand judgments

Theoretically accusative direct object is specific and nominative object is underspecified for specificity. However, in newspaper data accusative specific objects are rarely found. That is which we see in the table that for many transitive verbs the accusative object is not detected.

3.6 Usefulness of the SASU system

The SASU system is useful in three respects. Firstly, it can extract the subcategorization frames of verbs and classify them in transitive verbs, intransitive verbs and verbs of dative subject automatically. An example of each type of verbs is given in (28).

negatives; MCV=misclassified verbs

Subcat Frame	TP	FP	TN	FN	MCV	Accuracy
$\text{Subject}_{erg}\text{Object}_{acc}$	25	4	15	16	20	66.67 %
$\text{Subject}_{erg}\text{Object}_{nom}$	38	5	14	3	8	86.67 %
$\text{Subject}_{erg}\text{Clause}_{compl}$	8	8	43	1	9	85.00 %
$\text{Subject}_{nom}\text{Clause}_{compl}$	0	7	53	0	7	88.34 %
Subject_{nom}	14	5	37	4	9	85.00 %
Subject_{dat}	0	0	59	1	1	98.34 %
$\text{Subject}_{erg}\text{Obj}_{nom}\text{Obj}_{abl}$	12	3	39	6	9	85.00 %
$\text{Subject}_{erg}\text{Obj}_{nom}\text{Obj}_{loc.in}$	37	7	16	0	7	88.34 %
$\text{Subject}_{erg}\text{Obj}_{nom}\text{Obj}_{loc.on}$	0	0	57	3	3	95.00 %

Table 3.10: SCFs induced for 60 verbs

(28) a. علی نے کھانا کھایا.

ɑli=ne kʰana kʰa-ya
Ali=Erg meal.M eat-Perf.M
‘Ali ate the meal.’

b. ندا بازار گئی.

nɪda bazar gɑ-i
Nida.F market go-Perf.F.Sg
‘Nida went to the market.’

c. علی کو انعام ملا.

ɑli=ko ɪnam mɪl-a
Ali=Dat prize be-received-Perf.M.Sg
‘Ali got the prize.’

The other two uses of the system become available by extracting sentences corresponding to the unexpected CLCs and noting the context and grammatical functions in them. For one, different alternations of verbs are known. I have noted three basic types of alternations in Urdu verbs.

Intransitive/Transitive alternation:

Some verbs show an alternation between intransitive and transitive. For example, the verb *badal* 'change' shows such an alternation in Urdu (29).

(29) a. موسم مارچ میں بدلتا ہے.
mosɑm marc=mẽ bɑdɑl-ta hɛ
weather.M March=Loc.in change-Imperf.3Sg be.3Sg.Pres
'Weather changes in March.'

b. ندا نے کپڑے بدلے.
nɪda=ne kaṛe bɑdl-e
Nida.F=Erg clothes change-Perf.3Pl
'Nida changed the clothes.'

Some other verbs showing this alternation are *ga* 'sing' and $b^{h}ar$ 'fill'. A very few intransitive verbs sometimes take the object which is derived from the verb itself. An example of this type of verb is ʊṛ 'fly' which is illustrated in (30).

(30) a. پرندے اونچا اڑے.
pɑrɪnde ũca ʊṛ-e
bird.3Pl high fly-Perf.3Pl
'Birds flew high.'

b. پرندوں نے اونچی اڑان اڑی.
pɑrɪndõ=ne ũci ʊṛan ʊṛ-i
bird.3Pl=Erg hight flight fly-Perf.F.Sg
'Birds flew high.' lit: Bird flew the high flight.

Intransitive/Dative-Subject alternation: Another alternation which has been observed in Urdu verbs is between an intransitive verb and a dative subject verb. An example of a verb showing this alternation is $dɪk^{h}$ 'appear' (31).

(31) a. آج چاند نہیں دکھا.
aj cãd nɑhĩ dɪkʰ-a
today moon.M not appear-Perf.M.Sg
'The moon did not appear today.'

b. ندا کو جنگل میں ایک سانپ دکھا.
nɪda=ko jãgɑl=mẽ ek sãp dɪkʰ-a
Nida=Dat jungle.M=Loc.in one snake.M.3Sg appear-Perf.M.Sg
'Nida saw a snake in a jungle.'

Intransitive/Copula alternation:

Some intransitive verbs are alternated with the copular use. A typical example is the verb rɑh 'live/remain' (32).

(32) a. علی ملتان میں رہتا ہے.
ɑli mʊltan=mẽ rɑh-ta hɛ
Ali.M Multan=Loc.in live-Imperf.M.Sg be.3Sg.Pres
'Ali lives in Multan.'

b. ندا تین دن بیمار رہی.
nɪda tin dɪn bimar rɑh-i
Nida.F three day.M ill remain-Perf.F.Sg
'Nida remained ill for three days.'

Another verb *ho* 'be/beocme', which will be discussed in detail in Chapter 4, also shows this alternation.

Finally, sometimes unexpected CLCs of the verb under investigation are an indication that the verb forms complex predicates with other predicates. For, example, examining the sentences of unexpected CLCs of the verb *a* having the *ko* bit value equal to 1, the following complex predicates were identified.

(33) a. شرم آ ʃɑrɑm a 'to be shy'

b. رحم آ rɑhm a 'have mercy'

c. نظر آ nɑzɑr a ‘be-seen’

d. پیار آ pıyar a ‘feel love’

e. یاد آ yad a ‘recall’

All the complex predicates in (33) have the *ko* (dative) marked subject argument. So, the empirical observation of unexpected CLCs can help collect the lexicon of complex predicates in Urdu.

3.7 Limitations of the SASU system

The SASU system cannot recognise some token as a main verb with some lexical or structural cue. Verbs in the corpus are identified by comparing conjugation forms of verbs with all tokens of a sentence in the corpus. I have tested the SASU system only for simple predicates after having developed a repository of about 700 verbal roots and about 1200 stems in Urdu. However, once a repository of complex predicates is developed this system will also be able to extract their syntactic frames. In that case the conjugation of the light verb in a complex predicate would be matched with tokens of the corpus with an added rule that other predicate just occurs before the light verb.

Another limitation of the system is that it cannot extract infinitival arguments. Although infinitives themselves can be identified but in presence of some infinitive it is hard to identify whether some case phrase is of an infinitive or of a main verb. So, sentences of verbs with infinitival elements are not considered for extraction of CLCs of the verb. Still another limitation is that it cannot detect unmarked and genitive marked arguments and adjuncts of a verb from a raw corpus. As there is no clitic for an unmarked argument, so it cannot be identified from cues of just case clitics.

Although genitive case clitics exist, they are so abundantly used to mark modifiers/specifiers of head nouns that they cannot act as useful cues for genitive marked arguments/adjuncts of a verb. An example of a genitive marked adjunct and a genitive marked argument is illustrated in (34).

(34) a. علی کا بھائی رات کا آیا ہوا ہے.

ɑli=ka	b^{h}ai	rat=ka	ay-a	hu-a	hɛ
Ali.M=Gen	brother.M	night.F=Gen	come-Perf.M	be-Perf.M	be.3Sg.Pres

‘Ali’s brother is here since night.’

b. ندا نے علی کی بہن سے رابطے کا سوچا.

nɪda=ne	ɑli=ki	bɑhn=se	rabte=ka	soc-a
Nida.F=Erg	Ali.M=Gen.F	sister.F=Com	contact.M=Gen	think-Perf.M.Sg

‘Nida thought of contacting with Ali’s sister.’

There are two genitive marked elements in each instance of (34). The first genitive marked element in each instance is the modifier of the head noun. The second genitive marked element in (34a) is the adjunct of the verb *a* ‘come’ and the second genitive marked element in (34b) is the argument of the verb *soc* ‘think’.

Although I have not extracted adpositional participants of a verb from the the Urdu corpora, this can be made possible by incorporating more bits of adpositions in the information vector.

Subcategorization patterns of the verb *ho* ‘be’ due to its diversified syntacto-semantic behavior cannot be acquired by the developed SASU system. Multiple uses of this verb and its syntactic frames are empirically investigated in the next chapter.

3.8 Conclusion

The system/scheme presented in this chapter is very promising for inferring subcategorization frames of verbs in Urdu. As other South Asian languages like Saraiki, Sindhi, Balochi, Nepali etc., are very close to Urdu structurally, this scheme could also be applied to these languages. In principle, the approach is applicable to any case-rich language, whether of fixed word order or of free word order, if the nouns in the language are marked for case by separate lexical elements. The system does an extensive screening for the candidate sentences of the target verb, therefore if the experimentation is done on a sufficiently large and balanced Urdu corpus, the results are expected to improve further. As classification of German verbs was made based on frequencies of

subcategorization frames (Schulte im Walde 2006; Schulte im Walde & Brew 2002), the automatic classification of verbs in Urdu can be investigated directly from frequencies of CLCs.

Chapter 4

The verb *ho* 'be/become'

For encoding a given notion, different syntactic structures could be used in different languages or even within a single language. On the other hand it is also possible that different notions are encoded with similar syntactic structures. Urdu does both. In Urdu, the verb *ho* 'be/become' is especially interesting because the verb is not only used to express functions of copular predication, general stative expressions of existence and expressions of possession, but is also used as a dynamic verb and in adjectival participial expressions. In addition, *ho* 'be/become' is used as a tense auxiliary and as the verbal part (light verb) in Noun-Verb and Adjective-Verb complex predicates.

This verb has more conjugation forms than other basic verbs of Urdu. In Table 4.1, different inflectional forms of the verb *ho* 'be/become' are listed. The first three in this table are of infinitives, the next eight conjugation forms are non-aspectual and the last eight are aspectual ones. Its different conjugation forms are used in different contexts and constructions. Due to its diversified and not yet understood syntacto-semantic behavior it was not viable to extract its frames from the SASU system. So, they are empirically investigated in this chapter.

This chapter explores in detail the stative and dynamic uses of the verb *ho* and the number and type of arguments this verb can take. These points will be explored by investigating uses of infinitival *ho*, non-aspectual *ho* and aspectual *ho* in sections 4.1, 4.2 and 4.3, respectively. Syntactic tests which distinguish between stative and dynamic *ho* are provided in section 4.4. The verb *ho* as a light verb is discussed in section 4.5 and the semantic realizations of different constructions which make use of this verb are described in section 4.6. The characterizing participles, which also make use of the verb *ho*, are analyzed with respect to their possible subcategorization frames and the type of arguments they can modify in section 4.7. Section 4.8 concludes the chapter.

Form	Feature decomposition
ho-na	be.Inf.M.Sg
ho-ne	be.Inf.M.Pl
ho-ni	be.Inf.F
tʰa	be.M.3Sg.Past
tʰe	be.M.3Pl.Past
tʰi	be.F.3Sg.Past
tʰĩ	be.F.3Pl.Past
hɛ	be.2/3Sg.Pres
hɛ̃	be.Pl.Pres
ho	be.2.Pres OR be.2/3Sg.Subjn
hũ	be.1Sg.Pres OR be.1Sg/Pl.Subjn
hu-a	be.Perf.M.Sg
hu-e	be.Perf.M.Pl
hu-i	be.Perf.F.Sg
hu-ĩ	be.Perf.F.Pl
ho-ta	be.Imperf.M.Sg
ho-te	be.Imperf.M.Pl
ho-ti	be.Imperf.F.Sg
ho-tĩ	be.Imperf.F.Pl

Table 4.1: Conjugation forms of the verb *ho* 'be/become'

4.1 Infinitive *ho*

To illustrate the stative and the dynamic use of the verb *ho*, some example phrases with its infinitival form are given in (1). Although an infinitive also is inflected for number and gender agreement, only the default masculine singular form of the infinitive is used when the subject argument is marked for genitive case as in (1).

(1) a. بالوں کا لمبا ہونا

balõ=ka lɑmba ho-na
hair.3Pl=Gen.3Sg long.M.3Sg be-Inf.3Sg
'The hair being long' OR 'The hair becoming long'

b. بالوں کا لمبا ہو جانا.
balõ=ka lɑmba ho ja-na
hair.3Pl=Gen.3Sg long.M.3Sg be go-Inf.3Sg
'The hair becoming long'

c. تیل لگانے سے بالوں کا لمبا ہونا.
tel lɑga-ne=se balõ=ka lɑmba ho-na
oil.M.3Sg apply-Inf=Abl hair.3Pl=Gen.3Sg long.M.3Sg be-Inf.3Sg
'The hair becoming long by applying oil'

Both stative and dynamic readings are possible with the bare infinitive of *ho* (1a). However, if the light verb *ja* is used (1b), only dynamic meanings are possible. Stative meanings are also suppressed in the context of some adverbial phrases (1c). These adverbial phrases usually encode cause/agentive effect due to which dynamic readings become more plausible.

4.2 Non-aspectual *ho*

Non-aspectual *ho* is used as an existential verb, as a stative copula and as a tense auxiliary. The subjunctive forms of *ho*, when used with the future, yield a modal interpretation with stative or dynamic reading. A detailed discussion of these phenomena is provided in the following subsections.

4.2.1 The verb *ho* as an intransitive verb

The verb *ho* is used as a lexical verb in an 'existential' meaning. Some examples of existential meaning are given in (2). In these examples, the verb *ho* acts like an intransitive verb and takes only the subject argument.

(2) a. خدا ہے.
xʊda hɛ
God.M.Sg be.3Sg.Pres
'There is a God.'

b. آج کافی گرمی ہے.
aj kafi gɑrmi hɛ
today much hot.F.Sg be.3Sg.Pres
‘It is very hot today.’

c. کل عید تھی.
kɑl id tʰi
yesterday Eid.F be.F.Sg.Past
‘It was Eid yesterday.’

4.2.2 The verb *ho* as copula

Copular constructions can encode a variety of meanings (for details, see Declerck 1988; Hengeveld 1992). To encode meanings of identification and classification, there are four types of basic copula constructions (Curnow, 2000) i.e., (a) verbal copula, (b) particle copula, (c) zero copula and (d) inflectional copula. A language can have one or more copular constructions. The choice of construction depends upon factors such as tense and aspect, polarity, status of the clause as main or subordinate, person of the subject and the semantic relation expressed. Some examples of different copular constructions given by Curnow are in (3).

(3) a. ten chłopiec jest moim uczniem
this.Nom boy.Nom is my.Instr pupil.Instr
‘This boy is my pupil.’ (Polish, Comrie 1997:40)

b. is docht˙ir È
Cop doctor he.Acc
‘He is a doctor.’ (Modern Irish, Doherty 1996:2)

c. ni-ta:kat
1SgSubj-man
‘I am a man.’ (Pipil, Campbell 1985:54)

d. Sara mora
Sara teacher
‘Sara is a teacher.’ (Modern Hebrew, Junger 1981:122)

The verbal copula construction involves a verbal element, similar to the English *be*, as is shown in (3a). In the particle copula construction (3b) there is some word in addition to the copula subject and the copula complement. This word either does not inflect or inflects for categories different from those of the verb in the language. In the inflectional copula construction (3c), the copula complement itself is inflected. In the zero copula construction (3d), the copula subject and the copula complement are simply juxtaposed.

In Urdu a verbal copula construction is used to encode meanings of identification (4a) and class membership (4b).

(4) a. وہ لڑکا میرا بھائی ہے.

vʊh	lɑṛka	mera	b^hai	hɛ
3Sg.Nom	boy.M.3Sg	3Sg.Gen	brother	be.3Sg.Pres

‘That boy is my brother.’

b. علی ڈاکٹر ہے.

ɑli	ḍakṭɑr	hɛ
Ali.M.3Sg	doctor.M.3Sg	be.3Sg.Pres

‘Ali is a doctor.’

The instance in (4a) encodes the notion of identification of one entity with another entity and the instance in (4b) encodes membership of Ali in the class of people who are characterized as doctors. Many linguists (e.g. Declerck 1988) claim that copular sentences are basically either specificational or predicational, though these categories may be subdivided further. Here, (4a) is an instance of a specificational sentence in the sense that ‘that boy’ is specified by ‘my brother’ and (4b) is an instance of a predicational sentence as the title of ‘doctor’ is predicated of ‘Ali’. In Urdu, the verbal copula construction is used to encode all specificational and predicational sentences.

(5) a. علی ذہین ہے.

ɑli	zɑhin	hɛ
Ali.M	intellignt	be.3Sg.Pres

‘Ali is intelligent.’

b. ندا سولہ سال کی ہے۔
nıda solɑh sal=ki hɛ
Nida.F sixteen year=Gen be.3Sg.Pres
‘Nida is of 16 years.’

Some property can be predicated of the subject in ‘NP adjective *ho*’ construction. The property of being intelligent is attributed to ‘Ali’ in (5a) and the property of being 16 years old is attributed to ‘Nida’ in (5b). Attributive genitives like *solah sal=ki* ‘of 16 years’ show a distribution like that of adjectives in Urdu (Raza, 2010c). These are discussed in detail in Chapter 5.

(6) a. کتاب میز پر ہے۔
kıtab mez=pɑr hɛ
book.F.3Sg table.F.3Sg=Loc.on be.3Sg.Pres
‘The book is on the table.’

b. میز پر ایک کتاب ہے۔
mez=pɑr ek kıtab hɛ
table.F.3Sg=Loc.on one book.F.3Sg be.3Sg.Pres
‘There is one book on the table.’

The example in (6a) encodes a location relation and it is also a predicational sentence in the sense that the location, *mez=par* ‘on the table’ is predicated of the entity, *kıtab* ‘the book’.

$$\begin{bmatrix} \text{PRED} & \text{`ho}\left\langle(\uparrow \text{SUBJ})\ (\uparrow \text{PREDLINK})\right\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`kıtab'} \\ \text{PERS} & \text{3, NUM sg} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{PREDLINK} & \begin{bmatrix} \text{PRED} & \text{`mez'} \\ \text{PERS} & \text{3, NUM sg} \\ \text{CASE} & \text{loc} \end{bmatrix} \\ \text{TENSE} & \text{present} \end{bmatrix} \quad \begin{bmatrix} \text{PRED} & \text{`ho}\left\langle(\uparrow \text{SUBJ})\right\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`kıtab'} \\ \text{SPEC} & \text{`one'} \\ \text{PERS} & \text{3, NUM sg} \\ \text{CASE} & \text{nom} \end{bmatrix} \\ \text{ADJUNCT} & \left\{\begin{bmatrix} \text{PRED} & \text{`mez'} \\ \text{PERS} & \text{3, NUM sg} \\ \text{CASE} & \text{loc} \end{bmatrix}\right\} \\ \text{TENSE} & \text{present} \end{bmatrix}$$

Figure 4.1: f-sructure representations of (6a) and (6b)

I postulate two different analyses for the example in (6a) vs. the example in (6b). The location describes the relative position of some object/figure and is considered as PREDLINK (Butt *et al.*, 1999) in (6a). The example in (6b), however, can be

considered an intransitive lexical verb *ho* with the locative marked noun as an adjunct of the existential clause. In existential constructions, the locative phrase is usually in the subject position (Freeze, 1992).

It should be noted that the canonical order of different elements in copular constructions in Urdu is as following:

Copular-subject + Copular-complement + Copula

Although Urdu is generally considered a free word language, the order of different elements in copular construction does matter for the selection of the subject and the complement.

The sentence in (6b) would be uttered in a response to the question, *mez=par kıya hɛ?* 'What is on the table?' and the sentence in (6a) would be uttered in a response to the question *kıtab kahã hɛ?* 'Where is the book?'. In the answer of the first question the locative phrase can optionally be dropped and someone can just say *ek kıtab hɛ* 'It is a book.'. And, interestingly the first question can also be put in this way: *mez=par yıh kıya hɛ?* 'What is this on the table?'. The f-structures for (6a) and (6b) are given in Figure 4.1. The copular constructions in Urdu have been analyzed as a double-tier closed function, following Attia (2008). In double-tier analysis, instead of the predicate, the copula or some dummy copula (when it is not present) is considered as a clausal head. The predicate is considered a closed function, i.e. PREDLINK.

Sometimes, when the relation is one of 'integral' part to whole, the whole is marked either for genitive or for locative *mẽ* and appears in the subject (Masica 1991:359) position as in (7).

(7) کمرے کی / میں دو کھڑکیاں ہیں۔

kɑmre=ki/mẽ	do	kʰıṛkıyã	hɛ̃
room.M=Gen.F/Loc.in	two	window.F.3Pl	be.3Pl.Pres

'The room has two windows.'
='There are two windows in the room.'

(8) a. بلی میز کے نیچے ہے.

bɪlli mez=ke nice hɛ
cat.F.3Sg table.F.3Sg=Gen under be.3Sg.Pres
‘The cat is under the table.’

b. یہ کتاب قانون سے متعلق ہے.

yɪh kɪtab qanun=se mʊt’ɑlɪq hɛ
this book.F.3Sg law.M=Abl about be.3Sg.Pres
‘This book is about law.’

Another locative relation is encoded in (8a) by a spatial postpositional phrase that is acting as the PREDLINK of a copular construction. The non-spatial adpositional phrases can also act as PREDLINK as in (8b). In both instances of (8) adpositional phrases predicate ‘being under the table’ and ‘being about law’ of the entities ‘the cat’ and ‘this book’. Hence the former ones are acting as PREDLINKs and the latter ones as subjects.

4.2.2.1 Possession of abstract characteristics/properties

The locative marked nouns entailing the real location cannot be the subject of the construction, as was assumed in (6b). However, spatial case markers in Urdu are also used to mark the abstract location of some properties (Mohanan, 1994), in which case the abstract location can be the subject of the clause. For example, the clitic *par* is used to mark the abstract location of responsibility and burden (9) and the clitic *mẽ* is used to mark the abstract location of some inherent characteristics, such as *jʊr’ɑt* ‘courage’ and *mamta* ‘affection’, etc (10).

(9) a. علی پر بچوں کی تعلیم کی ذمّہ داری ہے.

ɑli=pɑr bɑccõ=ki tɑ’lim=ki zɪmmɑhdari hɛ
Ali=Loc.on child.M.3Pl=Gen.F education.F=Gen.F responsibility.F be.Pres.3Sg
‘Ali has responsibility of the education of the children.’

b. ندا پر بچوں کی پرورش کا بوجھ ہے۔

nɪda=pɑr	bɑccõ=ki	pɑrvɑrɪʃ=ka	bojʰ	hɛ
Nida=Loc.on	child.M.3Pl=Gen.F	upbringing.F=Gen.M	burden.F	be.Pres.3Sg

‘Nida has a burden of upbringing of the children.’

(10) a. علی میں بہت جرأت ہے۔

ɑli=mẽ	bɑhʊt	jʊr’ɑt	hɛ
Ali.M=Loc.in	much	courage	be.3Sg.Pres

‘Ali has a great deal of courage.’

b. ندا میں بڑی مامتا ہے۔

nɪda=mẽ	bɑṛi	mamta	hɛ
Nida.F=Loc.in	much	affection.F	be.3Sg.Pres

‘Nida has a great deal of affection.’

```
[ PRED      ‘ho⟨(↑ SUBJ) (↑ PREDLINK)⟩’
  SUBJ      [ PRED  ‘ɑli’
              PERS  3, CASE loc.in ]
  PREDLINK  [ PRED      ‘zɪmmɑhdari’
              ADJUNCT   { [ PRED     tɑ’lim
                            ADJUNCT  { [ PRED  bɑcca
                                         PERS  3, CASE gen ] }
                            PERS     3, NUM sg ] }
              PERS      3, NUM sg ]
  TENSE     present ]
```

Figure 4.2: f-structure representation of (9a)

The subject argument of the copular constructions in (9) and (10) is marked for locative case. The f-structure representation of (9a) is given in Figure 4.2. The instances in (10) have two alternate utterances each with the same truth-conditional meaning, for example, for (9a) two alternate utterances are given in (11) below.

(11) a. بچوں کی تعلیم کی ذمّہ داری علی پر/ کی ہے۔
[[bɑccõ=ki tɑ'lim]=ki zımmɑhdari] ɑli=pɑr/ki hɛ
[[child.Pl=Gen education.F]=Gen responsibility.F] Ali=Loc.on/Gen.F be.Pres.3Sg
'The responsibility of the education of the children is on Ali.'

b. بچوں کی تعلیم علی کی ذمّہ داری ہے۔
[bɑccõ=ki tɑ'lim] [ɑli=ki zımmɑhdari] hɛ
[bchild.M.3Pl=Gen.F education.F] [Ali=Gen.F responsibility.F] be.Pres.3Sg
'Education of the children is the responsibility of Ali.'

In the case of (11a), the abstract location would be the PREDLINK in the f-structure. However, (11b) is an instance of an NP+NP+*ho* construction. The abstract location of some inherent characteristic is marked either by the locative or the dative case as in (12). According to Mohanan (1994), *mẽ* 'in' expresses the location of a characteristic attribute that is relatively permanent while the dative *ko* encodes the abstract location of a temporary sate.

(12) a. علی میں کسی قسم کا لالچ نہیں ہے۔
ɑli=mẽ kısi qısm=ka lalɑc nɑhĩ hɛ
Ali.M=Loc.in any sort.F=Gen.M greed.M not be.3Sg.Pres
'Ali has no greed of any sort.'

b. ندا کو کسی چیز کا لالچ نہیں ہے۔
nıda=ko kısi ciz=ka lalɑc nɑhĩ hɛ
Nida.F=Dat any thing.F=Gen.M greed.M not be.3Sg.Pres
'Nida has no greed of anything.'

The relation between physical experience and the experiencer can also be encoded by copular construction in Urdu. In such constructions, the experiencer is mapped to the dative marked subject and the experience itself is mapped to the PREDLINK of the clause.

(13) a. علی کو بخار ہے۔
ɑli=ko bʊxar hɛ
Ali.M=Dat fever.M be.3Sg.Pres
'Ali suffers from fever.'

b. ندا کو کھانسی ہے۔
nɪda=ko kʰɑ̃si hɛ
Nida.F=Dat cough be.3Sg.Pres
'Nida suffers from cough.'

In (13), the diseases are related to the experiencer by copular *ho.* We see here that the subject which corresponds to the experiencer thematic role is marked for dative case. Let us see another instance of a dative marked element in the context of copular constructions.

(14) a. علی کو یہ قمیض ڈھیلی ہے۔
ɑli=ko yɪh qɑmiz dʰili hɛ
Ali.M=Dat this shirt.F.3Sg loose.F.3Sg be.3Sg.Pres
'This shirt is loose to Ali.'

b. ندا کو وہ پاجامہ تنگ ہے۔
nɪda=ko vʊh pajamɑh tɑ̃g hɛ
Nida.F=Dat that trouser.M.3Sg tight be.3Sg.Pres
'That trouser is tight to Nida.'

The dative NPs in (14) can be dropped and still the instances remain grammatical. Those datives in German and in some other languages that may be dropped without any syntactic and semantic residue are called free datives (Hole, 2009). Hole has given the following test to distinguish free datives from subcategorized datives in German:

(15) *A dative argument* D not dependent on a preposition is *free* in a simple positive declarative sentence S of German **iff**

(i) S without D is grammatical;

(ii) S without D does not entail that there is an individual

(α) which participates in the event described by S and

(β) which could be encoded as a dative argument. (Hole 2009:3)

The Urdu datives in (14) also act like the free datives of German. The free datives in these constructions come due to adjectival predicates. These constructions will be discussed more in section 5.3.

4.2.2.2 Possession of concrete objects

In English there is a verb *have* which is used to encode the possession relation. In Urdu, however, there is no such verb; instead the copular *ho* is used to encode the possession relation as in (16).

(16) a. علی کی ایک بیٹی ہے.[1]

ɑli=ki ek beṭi hɛ
Ali.M.3Sg=Gen.F one daughter.F.3Sg be.3Sg.Pres
'Ali has a daughter.'

b. ندا کے چار مکان ہیں.

nɪda=ke car mɑkan hɛ̃
Nida.F=Gen.3Pl four building.M be.3Pl.Pres
'Nida owns four buildings.'

The possessors in (16) are marked for genitive case and the copular verb agrees with the unmarked possessees. A verb in Urdu agrees with the unmarked argument (Mohanan, 1994); if there is no unmarked argument then the default masculine singular form is used. Mohanan (1994) has argued that the genitive marked element in these constructions acts as the subject.

There is still another copular construction in Urdu, which encodes the possession relation. In this construction the possessor is encoded as the complement of a proximal

[1]In Hindi, the genitive case clitic that agrees usually in number and gender with the possessum can be replaced with the frozen form *ke* in case of inalienable possession (Mohanan 1994:178). Masica reports that the invariant genitive is used to express the kinship relation (Masica 1991:360). However, in contemporary Urdu, the genitive case clitic always agrees with the possessum both in noun phrases and the copular construction.

spatial postposition *pas* ‘near’.[2] Examples are given in (17).

(17) a. علی کے پاس سونے کی انگوٹھی ہے.

ɑli=ke	pas	sone=ki	ɑ̃guʈʰi	hɛ
Ali.M.3Sg=Gen	near	gold.M=Gen.F	ring.F.3Sg	be.3Sg.Pres

‘Ali has a golden ring.’

b. ندا کے پاس پانچ مکان ہیں.

nɪda=ke	pas	pɑ̃c	mɑkan	hɛ̃
Nida.F=Gen.3Pl	near	five	building.M.3Sg	be.3Pl.Pres

‘Nida owns four buildings.’

Mohanan (1994) has again analyzed the nominal that bears the postposition *pas* in such constructions, as the subject. The f-strucure representation of (17b) is given in Figure 4.3. The postposition *pas* ‘near’ is used non-semantically. For full interpretation three features namely case, position-type and position form have also been used in the f-strucutre.

$$\begin{bmatrix} \text{PRED} & \text{‘ho}\left\langle(\uparrow \text{SUBJ})\ (\uparrow \text{PREDLINK})\right\rangle\text{’} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{‘nɪda’} \\ \text{PERS} & \text{3, CASE loc} \\ \text{PFORM} & \text{‘pas’} \\ \text{PTYPE} & \text{non-sem} \end{bmatrix} \\ \text{PREDLINK} & \begin{bmatrix} \text{PRED} & \text{‘mɑkan’} \\ \text{SPEC} & \text{pɑ̃c} \\ \text{PERS} & \text{3, CASE nom} \end{bmatrix} \\ \text{TENSE} & \text{present} \end{bmatrix}$$

Figure 4.3: f-structure representation of (17b)

Linguists have tried to distinguish the two possessive constructions; one with the genitive marked possessor and the other with the possessor as the complement of the postposition *pas* in Urdu/Hindi. McGregor (1972) holds that the genitive construction encodes the permanent possession relationship while the *pas* construction rather

[2] A similar construction is also found in Russian as is shown below:

U kati	(est’)	samovar
at Katia.Gen	(is)	samovar.Nom.Sg.M

‘Katia has a samovar.’ (Avgustinova & Uszkoreit, 2003)

expresses a 'contingent' possession. Pandharipande (1981), however, showed that McGregor's characterization is inadequate.

Pandharipande suggests that the genitive is used when there is emotional attachment or intimacy, for example, kinship or friendship and the *pas* construction is used to encode merely material possession. If there is ambiguity then either of the constructions can be selected. She argues that the choice between the two constructions depends merely on pragmatic factors. Recently, Sulger (2011) has differentiated the two constructions as individual level and stage level predication based on tests given by Carlson (1977), Diesing (1988) and Kratzer (1995).

(18) a. یہ قلم علی کا ہے.
yɪh qɑlɑm ɑli=ka hɛ
this pen.M.3Sg Ali.M=Gen.M3Sg be.3Sg.Pres
'This Pen is Ali's.'

b. میرا قلم ندا کے پاس ہے.
mera qɑlɑm nɪda=ke pas hɛ
my.Gen.M.3Sg pen.M3Sg Nida=Gen.Obl near be.3Pl.Pres
'My pen is with Nida.'

Some more examples of copular construction are given in (18). In these examples the possessums take the position of the subject. The f-structure of (18b) is given in Figure 4.4.

$$\begin{bmatrix} \text{PRED} & \text{`ho}\left\langle(\uparrow \text{SUBJ})\ (\uparrow \text{PREDLINK})\right\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`qɑlɑm'} \\ \text{SPEC} & \text{mera} \\ \text{PERS} & \text{3, CASE nom} \end{bmatrix} \\ \text{PREDLINK} & \begin{bmatrix} \text{PRED} & \text{`pas}\left\langle(\uparrow \text{OBJ})\right\rangle\text{'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`nɪda'} \\ \text{PERS} & \text{3, CASE gen} \end{bmatrix} \end{bmatrix} \\ \text{TENSE} & \text{present} \end{bmatrix}$$

Figure 4.4: f-structure representation of (18b)

I have asserted above that the order of elements matters for the determination of the subject in copular constructions. Mohanan (1994) has suggested some diagnostics for grammatical subjects in Hindi/Urdu, which are given in (19).

(19) a. The antecedent of *apna* must be a Logical-SuBJ or a SUBJ

b. The antecedent of a pronoun cannot be a c-commanding SUBJ of the minimal finite clause that contains the pronoun.

c. A nominal that can control a participial adjunct clause with an obligatory control site must be a SUBJ.

d. If an argument gaps or is gapped by a SUBJ, it must itself be a SUBJ (Mohanan 1994:135)

These tests can be used to determine subjects of simple and complex verbs in Urdu. However, these do not apply well in all cases of stative copular constructions. The word *apna* can appear in possessive copular constructions, not as a reflexive genitive, but in the meaning of 'own/personal'. The two uses of *apna* are illustrated in (20).

(20) a. علی نے ندا سے اپنا قلم مانگا.

ɑli=ne	nɪda=se	ɑpna	qɑlɑm	mɑ̃g-a
Ali.M=Erg	Nida=Abl	self.Gen.M.3Sg	pen.M.3Sg	ask-Perf.M.3Sg

'Ali$_j$ asked Nida for his$_j$ pen.'

b. علی نے ندا سے اس کا اپنا قلم مانگا.

ɑli=ne	nɪda=se	ʊs=ka	ɑpna	qɑlɑm	mɑ̃g-a
Ali.M=Erg	Nida=Abl	3Sg=Gen.M.3Sg	own.M.3Sg	pen.M.3Sg	ask-Perf.M.3Sg

'Ali asked Nida for her own pen.'

c. علی کے اپنے بیٹے نے اپنی دوست سے قلم مانگا.

ɑli=ke	ɑpne	beṭe=ne	ɑpni	dost=se	qɑlɑm	mɑ̃g-a
Ali.M=Gen	own.M.3Sg.Obl	son.M.3Sg.Obl=Erg	self.Gen.F	friend.F.3Sg=Abl	pen.M.3Sg	ask-Perf.M.3Sg

'Ali's own son$_j$ asked his$_j$ friend for the pen.'

(21) a. علی کی اپنی چار دکانیں ہیں۔

ɑli=ki ɑpni car dʊkanẽ hɛ̃
Ali.M=Gen.F own.F four shop.F.3Pl.Nom be.3Pl.Pres
'Ali has four shops of his own.'

b. یہ مکان ندا کا اپنا ہے۔

yɪh mɑkan nɪda=ka ɑpna hɛ
this house.M.3Sg.Nom Nida.F=Gen.M.3Sg own.M.3Sg be.3Sg.Pres
'This is Nida's own house.'

In (20a) *apna* is used as reflexive genitive pronoun and its antecedent is the subject of the clause. This *apna* was meant by Mohanan for diagnostic in (19a). However, in (20b) *apna* is used in the meaning of 'one's own' and its antecedent is not the subject of the clause. Both *apna* are present in (20c). In the possessive copular construction in (21) the morpheme *apna* is not used as genitive pronoun but it is used in the meaning of 'own'. The rest of the diagnostics also do not apply well in genitive copular constructions for the determination of the grammatical subject. The first participant should be considered as the subject in such constructions because this is the entity about which something is said/predicated.

4.2.2.3 Syntactic frames of stative copula

In light of the above discussion, the possible syntactic frames of stative copular *ho* are listed in Table 4.2.

This table shows a variety of syntactic phrases which can act as the two grammatical participants, the subject and the complement, of the stative copular verb in Urdu. The syntactic frames of dynamic copula will be listed after discussing the use of the subjunctive/future forms of *ho*.

4.2.3 The verb *ho* as modal

The verb *ho* is also used to construct modal meanings of 'obligation' and 'will', as shown in (22). The f-structure representation for (22b) following Butt & King (2005) is given in Figure 4.5.

Nr.	Subject	Complement (PREDLINK)	Example
1.	NP	NP	(4b)
2.	NP	AP	(5a)
3.	NP	KP_{Gen} (Attributive genitive)	(5b)
4.	NP (Possessum)	KP_{Gen} (Possessor)	(18a)
5.	NP	KP_{Loc} (Location)	(6a)
6.	NP	PP (Adpositional phrase)	(8)
7.	NP (Possessum)	PP_{pas} (Possessor)	(18b)
8.	NP marked with P_{pas} (Possessor)	NP (Possessum)	(17)
9.	KP_{Gen} (Possessor)	NP (Possessum)	(16)
10.	$KP_{Loc.in}$ (Whole)	NP (Part)	(7)
11.	$KP_{Loc.in}$ (Animate bearer)	NP (Abstract characteristic)	(10)
12.	$KP_{Loc.on}$ (Animate bearer)	NP (Responsibility/burden)	(9)
13.	NP (Responsibility/burden)	$KP_{Loc.on}$ (Animate bearer)	(11a)
14.	KP_{Dat} (Animate bearer)	NP (characteristic/disease)	(13)

Table 4.2: Syntactic frames of stative copular *ho*

(22) a. علی کو کام ختم کرنا تھا.

ɑli=ko kam xɑtɑm kɑr-na tʰa
Ali.M=Dat work.M finish do-Inf.M.3Sg be.M.3Sg.Past
'Ali had to finish the work.'

b. ندا کو جانا ہے.

nɪda=ko ja-na hɛ
Nida.F=Dat come-Inf.M.3Sg be.3Sg.Pres
'Nida has to go.'

c. ندا نے جانا ہے.

nɪda=ne a-na hɛ
Nida.F=Erg come-Inf.M.3Sg be.3Sg.Pres
'Nida is to go.'

$$\left[\begin{array}{ll}
\text{PRED} & \text{`ho}\left\langle\text{ja-na}\right\rangle(\uparrow \text{SUBJ})\text{'} \\
\text{XCOMP} & \left[\begin{array}{ll} \text{PRED} & \text{`ja-na}\left\langle(\uparrow \text{SUBJ})\right\rangle\text{'} \\ \text{SUBJ} & \boxed{7} \end{array}\right] \\
\text{SUBJ} & 7\left[\begin{array}{ll} \text{PRED} & \text{`nıda'} \\ \text{PERS} & \text{3, CASE dat} \end{array}\right] \\
\text{VTYPE} & \text{modal, TENSE present}
\end{array}\right]$$

Figure 4.5: f-structure representation of (22b)

(23) a. علی کام ختم کرنے کو تھا.

ɑli kam xɑtɑm kɑr-ne=ko tʰa
Ali.M work.M finish do-Inf=Dat be.M.3Sg.Past
'Ali was about to finish the work.'

b. ندا آنے کو ہے.

nıda a-ne=ko hɛ
Nida.F come-Inf.Obl=Dat be.3Sg.Pres
'Nida is about to come.'

c. ندا آنے والی ہے.

nıda a-ne=vali hɛ
Nida.F come-Inf.Obl=Val.F be.3Sg.Pres
'Nida is about to come.'

The verb *ho* is also used in another construction encoding an 'imminent' meaning in combination with dative marked infinitive (23a–b) or an infinitive marked with the *vala* clitic 'one' (23c).

4.2.4 The verb *ho* as tense auxiliary

There is often a formal identity between a copula verb and an auxiliary (Curnow, 2000). This is the case in Urdu as well. The non-aspectual copular forms of a stative *ho* are

also used as tense auxiliaries. Here are some examples.

(24) a. علی نے ایک کتاب لکھی تھی.
ɑli=ne ek kɪtab lɪkʰ-i tʰi
Ali.M.3Sg=Erg one book.F.3Sg write-Perf.M.3Sg be.F.3Sg.Past
'Ali had written a book.'

b. ندا کتاب پڑھتی ہے.
nɪda kɪtab pɑṛʰ-ti hɛ
Nida.F.3Sg book.F.3Sg read-Imperf.F.3Sg be.3Sg.Pres
'Nida reads a book.'

Because the perfect aspect by default includes the meaning of the past tense in it, the past tense auxiliary is usually not inserted to encode meanings in simple past. However, the past tense auxiliary is obligatory for encoding the meaning in the remote past (Butt & Rizvi, 2010).

4.2.5 Optionality of non-aspectual *ho*

Non-aspectual *ho* is sometimes suppressed when there is the negative element *nahĩ* 'not' and the tense is present. For example, the copular verb of present tense in the possession relation (25a–b) and the present tense auxiliary (25c) are optional when there is a negative element.

(25) a. ندا کی کوئی بہن نہیں (ہے).
nɪda=ki koi bɑhn nɑhĩ (hɛ)
Nida.F.3Sg=Gen.F any sister.F.3Sg not (be.3Sg.Pres)
'Nida has no sister.'

b. علی کے پاس کھانے کے لیے کچھ نہیں (ہے).
ɑli=ke pas kʰane=ke lɪye kʊcʰ nɑhĩ (hɛ)
Ali.M=Gen near eating=Gen for something not (be.3Sg.Pres)
'Ali has nothing to eat.'

c. علی دن کے وقت کچھ نہیں کھاتا (ہے).
ɑli din=ke vaqt kʊcʰ nɑhĩ kʰa-ta (hɛ)
Ali.M day.M=Gen time something not eat-Imperf.M (be.3Sg.Pres)
‘Ali eats nothing in day time.’

4.2.6 The verb *ho* with the future

The subjunctive *ho* in the future construction can have either a stative modal reading or a dynamic reading (26). The dynamic readings are usually possible in the presence of the *ja* light verb (26b) or with some adverbial phrase which encodes some cause effect (26c).

(26) a. ندا کے بال لمبے ہوں گے.
nɪda=ke bal lɑmbe hũ-ge
Nida.F.3Sg=Gen.M.3Pl hair.M long.M be.Subjn.Sg-Fut.M.Pl
‘Nida’s hair will be long.’
? ‘Nida’s hair will become long.’

b. ندا کے بال لمبے ہو جائیں گے.
nɪda=ke bal lɑmbe ho jaẽ-ge
Nida.F.3Sg=Gen.M.3Pl hair.M long.M be go-Subjn.3Pl-Fut.M.Pl
‘Nida’s hair will become long.’

c. تیل لگانے سے ندا کے بال لمبے ہوں گے.
[tel lɑga-ne=se] nɪda=ke bal lɑmbe hũ-ge
oil.M apply-Inf=Abl Nida.F=Gen.3Pl hair.M long.M be.Subjn.Sg-Fut.M.Pl
‘Nida’s hair will become long by using oil.’

(27) a. علی اس فلم کا ہیرو ہو گا.
ɑli ɪs fɪlm=ka hiro ho-ga
Ali.M this film=Gen.M hero be.Subjn.3Sg-Fut.M.3Sg
‘Ali will be the hero of this film.’

b. علی گلوکار ہو گا۔
ɑli gʊlukar ho-ga
Ali.M singer.M be.Subjn.3Sg-Fut.M.3Sg
* 'Ali will become the singer.'

c. علی گلوکار بنے گا۔
ɑli gʊlukar bɑne-ga
Ali.M singer.M become.Subjn.3Sg-Fut.M.3Sg
'Ali will become the singer.'

d. مونگ پھلی کھانے سے ندا کو کھانسی ہو گی۔
[mõg-pʰɑli kʰa-ne=se] nɪda=ko kʰãsi ho-gi
beans eat-Inf=Abl Nida.F=Dat cough.F be.Subjn.3Sg-Fut.F.3Sg
'Nida will become inflicted by cough by eating beans.

The syntactic frame NP+NP+ *ho* is possible only in case of stative *ho* as in (27a–b).[3] For the *NP become NP* construction in Urdu, another verb, *ban* 'become' is used as in (27c). However, the KP_{Dat}+NP+ *ho* frame is possible with dynamic *ho* (27d).

(28) a. دو سال بعد ندا ۱۸ سال کی ہو جائے گی۔
[do sal bɑ'd] nɪda 18 sal=ki ho jae-gi
two year after Nida.F 18 year.M=Gen.F be go.Subjn.3Sg-Fut.F.3Sg
'Nida will become 18 years old after two years.'

b. ایک اور دکان خریدنے سے علی کی پانچ دکانیں ہو جائیں گی۔
[ek ɔr dʊkkan xɑrid-ne=se] ɑli=ki pãc dʊkkanẽ
one more shop.F.3Sg buy-Inf.Obl=Abl Ali.M=Gen.F five shop.F.3Pl
ho jaẽ-gi
be go.Subjn.3Pl-Fut.F
'After purchasing one more shop Ali will possess five shops.'

[3]Some more modal readings are also possible: for (27a), 'Ali is probably the hero of this film.', and for (27b), 'Ali is probably the singer.' By adding *zarur* 'definitely' in (27a), another reading will be available, 'Ali must definitely be the hero of this film.'.

c. ایک اور قلم خریدنے سے ندا کے پاس دو قلم ہو جائیں گے.

[ek	ɔr	qɑlɑm	xɑrid-ne=se]	nɪda=ke	pas	do	qɑlɑm
one	more	pen.M.3Sg	buy-Inf.Obl=Abl	Nida.F=Gen.M	near	two	pen.M

ho jaẽ-ge
be go.Subjn.3Pl-Fut.M.3Pl
‘After purchasing one more pen Nida will possess two pens.’

d. شرط ہارنے پر یہ کار کسی اور کی ہو جائے گی.

[ʃɑrt	har-ne=par]	yɪh	kar	kɪsi	ɔr=ki	ho
bet.F	lose-Inf.Obl=Loc.on	this	car.F.3Sg	someone	else=Gen.F.3Sg	be

jae-gi
go.Subjn.3Sg-Fut.F.3Sg
‘On losing bet this car will be possesses by someone else.’

Nr.	Subject	Complement (PREDLINK)	Example
1.	NP	AP	(26b)
2.	NP	KP_{Gen} (Attributive genitive)	(28a)
3.	KP_{Gen} (Possessor)	NP (Possessum)	(28b)
4.	NP (Possessum)	KP_{Gen} (Possessor)	(28d)
5.	NP marked with P_{pas} (Possessor)	NP (Possessum)	(28c)
6.	KP_{Dat} (Animate bearer)	NP (Disease)	(27d)

Table 4.3: Syntactic frames of dynamic copular *ho*

Some more examples of dynamic *ho* are given in (28). The complement in (28a) is attributive genitive phrase, in (28b–c) is an NP possessum and in (28d) is the possessor. The possible syntactic frames of dynamic copular *ho* are listed in Table 4.3. The dynamic *ho* can also be an intransitive verb used in the meaning of ‘happen’ (29).

(29) a. پل کا افتتاح کل ہو گا.

pʊl=ka	ɪftɪtah	kɑl	ho	ga
bridge.F.3Sg=Gen.M	inauguration.M	tomorrow	be.Subjn.Sg	Fut.M.3Sg

‘The inauguration of the bridge will happen tomorrow.’

b. کل حکومت کے خلاف بڑا جلسہ ہو گا.
kɑl hɑkumɑt=Gen.M xılaf bɑṛa jɑlsɑh ho ga
tomorrow government.F=Gen against big procession.M be.Subjn.Sg Fut.M.3Sg
'A great procession against the government will happen tomorrow.'

We can also interpret examples in (29) with existential semantics, for example, (29a) can be paraphrased as: tomorrow there will exist an event of the bridge inauguration. However, the dynamic readings are also right, because the same constructions are also possible with the perfective and progressive aspects.

4.3 Aspectual *ho*

4.3.1 The imperfect form of *ho*

Imperfect forms generally are known to produce generic readings (Carlson & Pelletier, 1995). The imperfect form of the verb *ho* is used instead of the non-aspectual copular *ho* if some relation is generic in general.

(30) a. زرافے ک گردن لمبی ہوتی ہے.
zɑrafe=ki gɑrdɑn lɑmbi ho-ti hɛ
giraffe.M=Gen.F neck.F long.F.3Sg be-Imperf.F.3Sg be.3Sg.Pres
'A giraffe has a long neck.'

b. ایک یورو میں سو سینٹ ہوتے ہیں.
ek yoro=mẽ sɔ sanṭ ho-te hɛ̃
one Euro=Loc.in hundred cent be.Imoerf.Pl be.3Pl.Pres
'There are 100 cents in a Euro.'

c. مرغی کی دو ٹانگیں ہوتی ہیں.
mʊrɣi=ki do ṭãgẽ ho-ti hɛ̃
hen.F.3Sg=Gen.F.3Sg two leg.F.3Pl be-Imperf.F be.3Pl
'A hen has two legs.'

d. سانپ کی ٹانگیں نہیں ہوتیں.

sãp=ki	ṭãgẽ	nɑhĩ	ho-tĩ
snake.M.3Sg=Gen.F	leg.F.Pl	not	be-Imperf.F.3Pl

'A snake has no leg.'

A common noun in an Urdu sentence can be indefinite or specific depending on the context (Dayal, 1992). When the imperfect form of *ho* is used to encode some relation, the common noun in the subject position is interpreted as indefinite and represents the whole of its class. In (30) the nouns *zɑrafe* 'giraffe', *mʊrɣi* 'hen' and *sãp* 'snake' are indefinite nouns. Because each giraffe has a long neck, each Euro equals 100 cents, each hen has two legs and no snake has legs, these are all generic relations.

(31) a. رات کو شہر میں خاموشی ہوتی ہے.

rat=ko	ʃahr=mẽ	xamoʃi	ho-ti	hɛ
night.F=Temp	city.M=Loc.in	silence.F.Sg	be-Imperf.F.Sg	be.3Sg.Pres

'It is quiet at night in the city.'

b. سرما میں دن چھوٹے ہوتے ہیں.

sɑrma=mẽ	din	cʰoṭe	ho-te	hɛ̃
winter.M=Loc.in	day.M.3Pl	small.M.3Pl	be-Imperf.3Pl	be.3Pl.Pres

'Days are short in winter.'

c. جب امتحان نزدیک ہوتا ہے علی خوب پڑھتا ہے.

jɑb	ɪmtɪhan	nazdik	ho-ta	hɛ
when	examination.M.3Sg	near	be-Imperf.M.Sg	be.M.3Sg.Pres

ɑli	xub	pɑṛʰ-ta	hɛ
Ali.M.3Sg	much	study-Imperf.3Sg.M	be.3Sg.Pres

'Ali studies a lot whenever exams are near.'

d. زندگی وہاں ہوتی ہے جہاں پانی ہوتا ہے۔

zĩdgi	vɑhã	ho-ti	hɛ
life.F.3Sg	there	be-Imperf.F.3Sg	be.3Sg.Pres

jɑhã	pani	ho-ta	hɛ
where	water.M.3Sg	be-Imperf.M.3Sg	be.3Sg.Pres

'Life is there where water is.'

Some more examples of the use of the imperfect form of *ho* are given in (31). Because the city is quiet every night (31a), days are short every winter (31b) and Ali studies well each time when exams are near (31c), therefore these are notions of habitual/generic sense in specific intervals or at specific instants and are encoded by using imperfect *ho* rather than a simple non-aspectual copula. The example in (31d) also encodes the generic relation of water to life. From now onwards, for such generics, I would use the term 'contextual generics'.

In all instances, so far, the insertion of imperfect *ho* is obligatory to encode the generic meanings. Some examples are given in (32) where the imperfect *ho* can be inserted optionally.

(32) a. یہاں ایک درخت (ہوتا/ ہوا کرتا) تھا۔[4]

yɑhã	ek	daraxt	(ho-ta/hu-a kɑr-ta)	t^{h}a
here	one	tree.M.3Sg	(be-Imperf.M.Sg/be-Perf.M do-Imperf.M)	be.3Sg.Past

'There has been a tree here.'

b. علی آج کل انگلینڈ میں (ہوتا) ہے۔

ɑli	aj-kal	ĩglẽd=mẽ	(ho-ta)	hɛ
Ali.M.3Sg	now-a-days	England=Loc.in	(be-Imperf.M.Sg)	be.3Sg.Pres

'These days, Ali is in England.'

The imperfect form of *ho* is stylistically used in instances of (32). However, here also the presence of imperfect *ho* is explained along generic/habitual lines. In (32a) the being of a tree at some place is generic in some stretch of time in past and in (32b) the presence of Ali in England is generic in context of these days.

[4]Sometimes, the verbal combination *hu-a kar* is used to express the habitual notion.

Examples of imperfect *ho* shown so far have stative readings, where the imperfect form is used to encode defining properties and generic notions. The imperfect form of copular *ho* can also have a dynamic reading in an NP+AP+*ho* construction, as shown in (33).

(33) a. تیل لگانے سے بال لمبے ہوتے ہیں.
tel lɑga-ne=se bal lɑmbe ho-te hɛ̃
oil.M use-Inf=Abl hair long be-Imperf.3Pl be.M.3Pl.Pres
'Hair become long by using oil.'

b. سگریٹ نوشی سے پھیپھڑے خراب ہوتے ہیں.
sɪgret-noʃi=se pʰepʰr̤e xɑrɑb ho-te hɛ̃
cigarette-inhaling=Abl lung.M.3Pl imparied be-Imperf.3Pl be.3Pl.Pres
'Lungs become impaired by smoking.'

We have seen both in case of future and in case of imperfect form of *ho* that dynamic readings of copular verb are only possible if there is the use of the light verb *ja* or some adverbial construction is present.

4.3.2 The perfect form of *ho*

With perfect *ho*, generally dynamic readings are possible. In (34) are two examples where the verb *ho* is used with the meaning of 'become'.

(34) a. علی بہت خوش ہوا.
ɑli bɑhʊt xʊʃ hu-a
Ali.M much happy be-Perf.M.3Sg
'Ali became much happy.'

b. ندا سولہ سال کی ہوئی.
nɪda solɑh sal=ki hu-i
Nida.F.3Sg sixteen year=Gen.F.3Sg be-Perf.F.3Sg
'Nida became 16 years old.'

(35) a. پل کا افتتاح کل ہوا.

pʊl=ka ɪftɪtah kɑl hu-a
bridge.F.3Sg=Gen.M inauguration.M yesterday be-Perf.M.3Sg
'The inauguration of the bridge happened yesterday.'

b. کل حکومت کے خلاف بڑا جلسہ ہوا.

kɑl hɑkumɑt=ke xɪlaf bɑɽa jɑlsɑh hu-a
tomorrow government.F=Gen.Obl against big procession.M be-Perf.M.3Sg
'A great procession against the government happened yesterday.'

c. پہلی عالمی جنگ ۱۹۱۴ میں ہوئی.

pɑhli almi jɑ̃g 1914=mẽ hu-i
first of-world war 1914=Loc.in be-Perf.F.3sg
'The First World War occurred in 1914.'

d. سیلاب سے کافی تباہی ہوئی.

sɛlab=se kafi tɑbahi hu-i
flood.M.Sg=Abl enough destruction.F.Sg be-Perf.F.3Sg
'Much destruction occurred due to floods.'

In (35) the verb *ho* is used intransitively with the meaning of 'happen/occur/take-place'.

4.3.2.1 Perfect *ho* with present and future interpretation

(36) a. اچھا ہوا کہ ندا خریداری کے لیے بازار چلی گئی.

ɑccʰa hu-a kɪh nɪda xɑridari=ke lɪye bazar
good.M.Sg be-Perf.M.Sg that Nida.F shopping.F=Gen.Obl for market.M
cɑl-i gɑ-i
walk-Perf.F.Sg go-Perf.F.Sg
'It is good that Nida has gone to the market for shopping.'
'It turned/happened good that Nida has gone to the market for shopping.'

b. اگر جوتے مہنگے ہوئے تو ندا نہیں خریدے گی.

ɑgɑr jute mɑh̃ge hu-e to nɪda nɑhĩ xɑrid-e-gi
if shoes expensive be-Perf.Pl then Nida.F not purchase-Subjn-Fut.F
‘Nida will not purchase the shoes if they are (would be) costly.’

In (36a) *acc*^{h}*a hu-a* can have stative reading ‘it is good’ in present tense or the canonical dynamic reading ‘it turned/happened good’. In case of an if clause for the non true future tense the perfect form of the verb is used in Urdu as in (36b).

4.3.2.2 Perfect *ho* as an emphasis on being

(37) a. ہم ہوئے کہ تم ہوئے کہ میر ہوئے

اس کی زلفوں کے سب اسیر ہوئے

hɑm hu-e kɪh tʊm hu-e kɪh mir hu-e
1Pl be-Perf.Pl or 2Pl be-Perf.Pl or Mir be-Perf.Pl

ʊs=ki zʊlfõ=ke sɑb ɑsir hu-e
3Sg=Gen.F hair.F=Gen.M all prisoner be-Perf.Pl

‘Whether we, or you or Mir – all are bewitched by her hair’ (Mir Taqi Mir)

b. میرے حال پہ حیرت کیسی درد کے تنہا موسم میں

پتھر بھی رو پڑتے ہیں انسان تو پھر انسان ہوا

mere hal=pɪh hɛrɑt kɛsi dɑrd=ke tɑnha mosɑm=mẽ
my status.M=Loc.on wonder which pain=Gen lonely weather.M=Loc.in

pɑt^{h}ɑr b^{h}i ro pɑṛ-te hɛ̃ ɪnsan to p^{h}ɪr ɪnsan hu-a
stone.M also weep lie-Imperf.3Pl be.Pres man Emp then man be-Perf

‘Why is this wonder on my affairs, in the lonely weather of pain;
stones even begin to weep (for me), a man, though, is a man’ (Mohsin Naqvi)

c. پھول بھی ہوں درمیاں تو فاصلے ہوئے

p^{h}ʊl b^{h}i hũ dɑrmɪyã to fasle hu-e
flower.M even be.Subjn.Pl between then distance.M.Pl be-Perf.Pl
‘Even if flowers are in between, there are distances.’ (A line of a song)

Sometimes, the perfect form of *ho* is used to emphasize the being/participation of some subject in some context as is shown in different instances of (37).

In this section it was shown that aspectual forms of the verb *ho* are generally used to encode dynamic meanings. The imperfect form of the verb, however, is also used to encode generic notions. In the next section some tests are listed which distinguish between stative and dynamic uses of the verb.

4.4 Classification and distinction of the verb *ho* uses

Vendler (1968) divided verbs into four classes based on his Aktionsart schemata. However, basically all verbs can be classified into two major classes; stative verbs and dynamic verbs. Stative verbs involve no change, dynamic verbs do (Dowty, 1979).

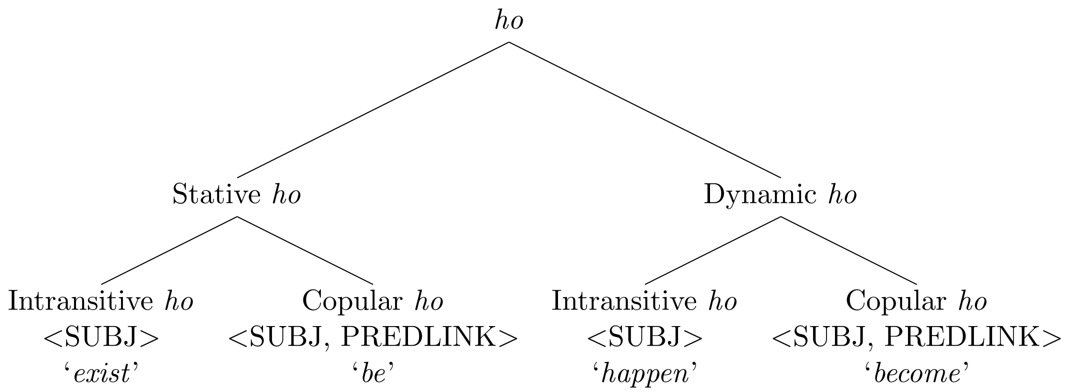


Figure 4.6: Classification of the verb *ho*

The verb *ho* is used both as a stative and as a dynamic verb in Urdu as shown in Figure 4.6. There are some syntactic tests that distinguish the stative use of *ho* from the dynamic use of *ho*. These are given in the following subsections.

4.4.1 Aspectual distinction

Stative and dynamic *ho* differ in having different aspects (Satyanarayana, 1982). The stative *ho* does not take perfect aspect (38a) while the dynamic *ho* can take perfect aspect (38b).

(38) a. کتاب میز پر ہوئی ہے۔
*kɪtab mez=pɑr hu-i hɛ
book.F.3Sg table.F.3Sg=Loc.on be-Perf.F be.3Sg.Pres
‘The book has been on the table.’ (Stative *ho*)

b. سیلاب سے کافی تباہی ہوئی ہے۔
sɛlab=se kafi tɑbahi hu-i hɛ
flood.M.Sg=Abl enough destruction.F.Sg be-Perf.F.3Sg be.3Sg.Pres
‘Enough destruction has occurred due to floods.’ (Dynamic *ho*)

Furthermore, stative *ho* cannot take progressive aspect (39a) while dynamic *ho* can take progressive aspect (39b).

(39) a. کتاب میز پر ہو رہی ہے۔
*kɪtab mez=pɑr ho rɑh-i hɛ
book.F.3Sg table.F.3Sg=Loc.on be Prog-F be.3Sg.Pres
‘The book is being on the table.’ (Stative *ho*)

b. سیلاب سے کافی تباہی ہو رہی ہے۔
sɛlab=se kafi tɑbahi ho rɑh-i hɛ
flood.M.Sg=Abl enough destruction.F.Sg be Prog-F be.3Sg.Pres
‘Much destruction is occurring due to floods.’ (Dynamic *ho*)

(40) a. علی تین دن بیمار تھا۔
ɑli tin dɪn bimar t^{h}a
Ali.M three day.M ill be.3Sg.Past
‘Ali was ill for three days.’ (Stative *ho*)

b. علی تین دن بیمار رہا۔
ɑli tin dɪn bimar rɑh-a
Ali.M three day.M ill remain-Perf.3Sg
‘Ali was ill for three days.’ (Stative *rah*)

It should be noted that another copular verb *rah* ‘remain’ can take perfect aspect to encode the stative meanings in a stretch of time in past (40b).

4.4.2 Light verb distinction

The *ja* 'go' light verb cannot combine with the stative *ho* (41a), however, its combination with the dynamic *ho* is possible (41b).

(41) a. کتاب میز پر ہو جاتی ہے.

*kɪtab mez=pɑr ho ja-ti hɛ
book.F.3Sg table.F.3Sg=Loc.on be go-Imperf.F.3Sg be.3Sg.Pres
'The book is on the table.' (Stative *ho*)

b. سیلاب سے کافی تباہی ہو جاتی ہے.

sɛlab=se kafi tɑbahi ho ja-ti hɛ
flood.M.Sg=Abl enough destruction.F.Sg be go-Perf.F be.3Sg.Pres
'Much destruction occurs due to flood.' (Dynamic *ho*)

Likewise another light verb *le* 'take' combines with dynamic *ho* but not with stative *ho*.

4.4.3 Auxiliary distinction

The construction formed with the combination of the default masculine perfect form of the main verb and the imperfect form of the auxiliary *cah* 'wish/want' is possible with dynamic *ho* as in (42b) but not with stative *ho* (Raza, 2010b) as in (42a).

(42) a. کتاب میز پر ہوا چاہتی ہے.

*kɪtab mez=pɑr hu-a cah-ti hɛ
book.F.3Sg table.F.3Sg=Loc.on be-Perf want-Imperf.F.3Sg be.3Sg.Pres
'The book is going to be on the table.' (Stative *ho*)

b. سیلاب سے کافی تباہی ہوا چاہتی ہے.

sɛlab=se kafi tɑbahi hu-a cah-ti hɛ
flood.M.Sg=Abl enough destruction.F.Sg be-Perf want-Perf.F be.3Sg.Pres
'Much destruction is going to occur due to floods.' (Dynamic *ho*)

4.4.4 Complement distinction

Not all syntactic frames of stative copular *ho* are possible for dynamic copular *ho*. For example, dynamic copular *ho* cannot take the NP complement (43b).

(43) a. علی ڈاکٹر ہے.
ɑli ḍakṭɑr hɛ
Ali.M.Nom doctor.M.3Sg be.3Sg.Pres
'Ali is a doctor.' (Stative *ho*)

b. علی ڈاکٹر ہوا.
* ɑli ḍakṭɑr hu-a
Ali.M.Nom doctor.M.3Sg be-Perf.M.Sg
'Ali became a doctor.' (Dynamic *ho*)

Different verb forms and their general use in a stative/dynamic sense are lsummarized in Table 4.4.

Verb forms	**Stative use**	**Dynamic use**
Infinitival	yes	yes
Subjunctive	yes	yes
Non-aspectual non-subjunctive	yes	no
Imperfect-aspectual	yes	yes
Perfect-aspectual	no	yes

Table 4.4: Uses of different forms of the verb *ho* 'be/become'

The imperfect aspectual forms in its stative use encode the contextual generic relations, the absolute generic relations and also the contextual generic existentials.

4.5 The verb *ho* as a light verb

In Urdu there are many Noun+Verb, Adjective+Verb and Verb+Verb complex predicates (Butt, 1995; Mohanan, 1994). The first part (noun, adjective or verb) in these predicates actually carries the core meaning and the second part (verb) is inflected for agreement and is called the light verb. There are more than a dozen verbs in Urdu which act as light verbs to form complex predicates. The verb *ho* is one of these verbs. Some examples of complex predicates with the light verb *ho* are given in (44).

(44) a. یاد ہو *yad ho*, lit: memory be, 'to remember'

b. شروع ہو *ʃʊru ho*, lit: start occur, 'to begin'

c. تعمیر ہو *ta'mir ho*, lit: building become, 'to become built'

d. حاصل ہو *hasıl ho*, lit: received be/become, 'to be/become received'

e. بھروسہ ہو *bʰarosah ho*, lit: trust be/become, 'to trust'

f. پیار ہو *pıyar ho*, lit: love be/become, 'to be/fall in love'

g. معلوم ہو *xabar ho*, 'to know/come to know'

h. خبر ہو *xabar ho*, lit: news be/become, 'to know/come to know'

As it was already shown that the verb *ho* can act both as a stative and as a dynamic verb in different constructions, the complex predicates formed with this verb can also be stative or dynamic. If the light verb *ho* is inflected with non-aspectual morphology of *ho* then we can say that it is a stative predicate (45), else it is a dynamic predicate (46). Sometimes both cases are possible (47).

(45) a. علی کو ندا کا انتظار تھا.
ɑli=ko nıda=ka ıntızar tʰa
Ali.M.3Sg=Dat Nida.F=Gen.M wait be.3Sg.Past
'Ali was waiting for Nida.'

b. علی کو ندا کا انتظار ہوا.
*ɑli=ko nıda=ka ıntızar hu-a
Ali.M.3Sg=Dat Nida.F=Gen.M wait be-Perf.M.Sg
'Ali was waiting for Nida.'

(46) a. مکان ۱۹۴۰ میں فروخت تھا.
*mɑkan 1940=mẽ fɑroxt tʰa
house.M 1940=Loc.in selling be.3Sg.Past
‘The house was sold in 1940.’

b. مکان ۱۹۴۰ میں فروخت ہوا.
mɑkan 1940=mẽ fɑroxt hu-a
house.M 1940=Loc.in selling be-Perf.M.Sg
‘The house was sold in 1940.’

(47) a. علی کو خبر ہے کہ ...
ɑli=ko xɑbɑr hɛ [kɪh ...
Ali.M.3Sg=Dat news be.3Sg.Pres [that ...
‘Ali knows [that ...’

b. علی کو خبر ہوئی کہ ...
ɑli=ko xɑbɑr hu-i [kɪh ...
Ali.M.3Sg=Dat news be-Perf.F.3Sg [that ...
‘Ali came to know [that ...’

In (47a) the stative meaning of ‘know’ is realized and in (47b) the dynamic meaning ‘come to know’ is realized. It should be noted that there are dative subjects also in copular constructions. In copular constructions there are only two arguments that is the subject and the complement. The main predicate in complex predicates of the verb *ho* introduces another argument. The that-clause argument in (47b) is introduced by the predicate *xabar* ‘news’. The main predicate of complex predicate cannot be coordinated with other predicate in contrast with the predicate in copular construction.

4.6 Various other uses of *ho*

Some forms of the verb *ho* are used to form constructions with different shades of meanings. The perfect participles of *ho* are used with participles of main verbs to encode

subjunctive meanings of desire (48b1–b2) and the subjunctive forms of *ho* together with the auxiliary *ga* are used with participles of main verbs to encode subjunctive meanings of possibility/likelihood (48c1–c2).

(48) a1. علی آیا.
ɑli a-ya
Ali.M come-Perf.M
'Ali came.'

a2. علی آتا.
ɑli a-ta
Ali.M come-Imperf.M
'Would that, Ali came.'

b1. علی آیا ہوتا.
ɑli a-ya ho-ta
Ali.M come-Perf.M be-Imperf.M
'Would that, Ali came.'

b2. علی آتا ہوتا.
ɑli a-ta ho-ta
Ali.M come-Imperf.M be-Imperf.M
'Would that, Ali used to come.'

c1. علی آیا ہو گا.
ɑli a-ya ho ga
Ali.M come-Perf.M be.Subjn Fut.M
'Ali may have come.'
'Ali is likely to have come.'

c2. علی آتا ہو گا.
ɑli a-ta ho ga
Ali.M come-Imperf.M be Fut.M
'Might be, Ali used to come.'
'Ali is probably about to come.'

The perfect participles of *ho* are also used to construct subordinate clauses. The action/state of such subordinate clauses and action of main verbs both are simultaneous. Consider, for example, the sentences in (49).

(49) a. ندا نے خط لکھا.
nɪda=ne xɑt lɪkʰ-a
Nida.F.3Sg=Erg letter.M.3Sg write-Perf.M.3Sg
'Nida wrote a letter.'

b. ندا نے مسکراتے ہوئے خط لکھا.
nɪda=ne [mʊskɑra-te hu-e] xɑt lɪkʰ-a
Nida.F.3Sg=Erg smile-Imperf.Obl be-Perf.Obl letter.M.3Sg write-Perf.M.3Sg
'Nida wrote a letter while smiling.'

c. Nida shrieb einen Brief [lächelnd] .
Nida.F.3Sg write.Past a.M.Sg.Acc letter.M.Sg smiling
'Nida wrote a letter while smiling.' (German)

d. ندا نے سیب کھاتے ہوئے خط لکھا۔
nɪda=ne [seb kʰa-te hu-e]
Nida.F.3Sg=Erg apple.M.3Sg eat-Imperf.Obl be-Perf.Obl
xɑt lɪkʰ-a
letter.M.3Sg write-Perf.M
'Nida wrote a letter while eating an apple.'

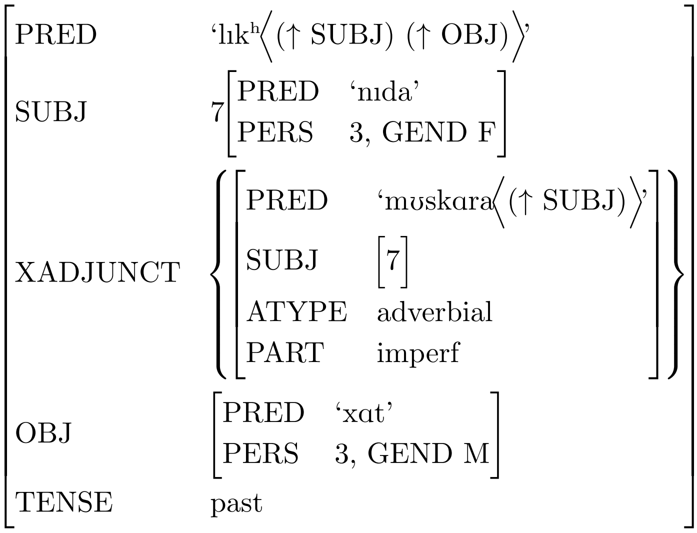


Figure 4.7: f-structure representation of (49b)

The example sentence in (49a) illustrates a simple sentence. The example sentence in (49b) adds a subordinate clause *mʊskarate hue* 'while smiling'. The actions of writing and smiling are simultaneous in (49b). The f-structure representation of (49b) is shown in Figure 4.7. German also has a similar construction, as shown in (49c). Another example with a transitive verb in the subordinate clause is shown in (49d). The subject of the verb in the subordinate clause is controlled by the subject of the main clause.[5]

[5] However, with some verbs like *ho* 'be' and *cah* 'wish', the subordinate clause has its own subject marked for genitive case as in (50).

(50) a. استاد کے ہوتے ہوئے بچے شور کر رہے تھے۔
[ʊstad=ke ho-te hu-e] bɑcce ʃor kɑr rɑhe tʰe
teacher.M=Gen be-Imperf.Obl be-Perf child.Pl noise do Prog.M be.3Pl.Past
'Children were making a noise in the presence of the teacher.'

b. عوام کے نہ چاہتے ہوئے فوج نے ملک پہ حکومت کی۔
[ɑvam=ke nɑh cah-te hu-e] fɔj=ne mʊlk=pɪh hɑkumɑt ki
public.F=Gen not want-Imperf be-Perf army.F=Erg country.M=Loc.on rule.F do-Perf.F
'The army unwanted by the public ruled over the country.'

(51) a. ندا نے یہ بیٹھتے ہوئے کہا۔

nıda=ne yıh [bɛṭʰ-te hu-e] kɑh-a

Nida.F.3Sg=Erg this sit-Imperf.Obl be-Perf.Obl say-Perf.F.3Sg

'Nida said this while seating.'

b. ندا نے یہ بیٹھے ہوئے کہا۔

nıda=ne yıh [bɛṭʰ-e hu-e] kɑh-a

Nida.F.3Sg=Erg this sit-Perf.Obl be-Perf.Obl say-Perf.F.3Sg

'Nida said this while having seated.'

(52) a. ندا نے یہ چلتے ہوئے کہا۔

nıda=ne yıh [cɑl-te hu-e] kɑh-a

Nida.F.3Sg=Erg this walk-Imperf.Obl be-Perf.Obl say-Perf.F.3Sg

'Nida said this while walking.'

b. ندا نے یہ چلے ہوئے کہا۔

*nıda=ne yıh [cɑl-e hu-e] kɑh-a

Nida.F.3Sg=Erg this walk-Perf.Obl be-Perf.Obl say-Perf.F.3Sg

'Nida said this while having walked.'

(53) a. ندا لیٹی (ہوئی) کتاب پڑھ رہی ہے۔

nıda [leṭ-i (hu-i)] kıtab pɑṛh rɑh-i hɛ

Nda.F lie-Perf.F be-Perf.F book.F read Prog-F be.3Sg.Pres

'Nida is reading a book while lying.'

b. ڈاکیا اپنے کندھے پر تھیلا لٹکائے (ہوئے) پھر رہا ہے۔

ḍakıya [ɑpne kɑndʰe=pɑr tʰɛla lɑṭka-e (hu-e)]

postman Gen.Refl shoulder.M=Loc.on bag.M hang-Perf.Obl (be-Perf.Obl)

pʰır rɑh-a hɛ

wander Prog-M be.3Sg.Pres

'A postman is wandering while hanging a bag on his shoulder.'

The perfect participle of *ho* is used with the perfect participles of only stative verbs (not action verbs) in subordinate clauses. The verb *bɛṭ*^{h} 'sit' in Urdu acts as a stative as well as a dynamic verb. In (51a) it is acting as a dynamic verb where it means 'Nida said this while taking her seat'. In (51b) the perfect participle of this verb is combined with the perfect participle of *ho* where it is interpreted as 'Nida said this while she has already taken the seat'. In (52b), we see that the perfect participle of the verb *cal* 'walk' cannot combine with *hu-e* because the verb *cal* 'walk' is a dynamic verb. So, we can distinguish between postural stative verbs and dynamic verbs in Urdu with this use of *ho*.

When the subject of the main clause is unmarked, the perfect participle of an intransitive stative verb of the subordinate clause can be in agreement with the subject and the perfect *ho* is optional (53a). However, the perfect participle of transitive stative verb of subordinate clause is obligatorily in the oblique form (53b).

In this section use of the verb *ho* in various subjunctive clauses and subordinate clauses was described. In the next section use of this verb in participial adjectives is explored.

4.7 Characterizing participles

The perfect form of *ho* is used to form two kinds of participles which characterize/modify a noun which logically is some argument of the verb. These participles, I call characterizing participles. In this section the syntactic frames of such participles, the arguments they modify and their syntactic distribution are explored.

4.7.1 Concomitant participles

Concomitant or present participles are formed by adding the perfect form of *ho* to the imperfect form of the verb. Sometimes, stylistically the mere imperfect form of the main verb is used as a characterizing participle. Consider some examples from Urdu and German in (54).

(54) a1. اڑتے (ہوئے) پرندے
ʊṛ-te (hu-e) pɑrɪnde
fly-Imperf.M (be-Perf.M) bird.M
'Flying birds'

a2. Die fliegenden Vögel
the fly.PresPart bird.F.Pl.Nom
'Flying birds'

b1. پٹتا (ہوا) مجرم
pɪṭ-ta (hu-a) mʊjrɪm
be-beaten-Imperf.M (be-Perf.M) criminal.M
'The criminal being beaten'

b2. No equivalent intransitive verb

c1. گرتا (ہوا) سیب
gɪr-ta (hu-a) seb
fall-Imperf.M (be-Perf.M) apple.M
'A falling apple'

c2. Ein fallender Apfel
A fall.PresPart apple.M.Sg.Nom
'A falling apple'

d1. بچے کو چومتی (ہوئی) ماں
bɑcce=ko cum-ti (hu-i) mã
baby.Sg=Acc kiss-Perf.F (be-Perf.F) mother.F
'The mother, kissing the baby'

d2. Die das Baby küssende Mutter
the.F.Nom the.Neut.Acc baby.Neut.Sg kiss.PresPart.F mother.F
'The mother, kissing the baby'

e1. کتاب پڑھتی (ہوئی) لڑکی
kɪtab pɑṛ^{h}-ti (hu-i) lɑṛki
book.F.3Sg read-Imperf.F (be-Perf.F) girl.F.3Sg
'The girl, reading the book'

e2. Das das Buch lesende Mädchen
the.Neut.Nom the.Neut.Acc book.Neut read.PresPart girl.Neut.Sg
'The girl, reading the book'

f1. لڑکی کو پھول دیتا (ہوا) لڑکا
lɑṛki=ko p^{h}ul de-ta (hu-a) lɑṛka
girl.F=Dat flower.M give-Imperf.M (be-Perf.M.Sg) boy.M.Sg
'The boy giving a flower to a girl'

f2. Der dem Mädchen die Blume gebende Junge
the.M.Nom the.Dat girl.Neut.Sg the.F flower.F give.PresPart boy.M.Sg
'The boy, giving a flower to a girl'

Like adjectives, characterizing present participles in Urdu and German also inflect to agree with the head nouns they modify, as shown in (54). The attributive distribution of characterizing present participles in Urdu, German and English is given in Table 4.5. Patient oriented intransitive verbs like *pɪṭ* 'be-beaten' have no intransitive equivalents

	Intransitive Verb			(Di)Transitive Verb		
Orientation	Agent (unerg)	Patient	Theme (unacc)	Agent	Patient	Theme
Urdu	yes	yes	yes	yes	no	no
English	yes	N.A	yes	no	no	no
German	yes	N.A	yes	yes	no	no

Table 4.5: Attributive use of characterizing present participles

in German and English. So, patient oriented intransitive characterizing participles are not found in German and English. Characterizing participles of transitive verbs can only modify the agent argument, all other arguments appear as the arguments of participles.

In German, past participles with an adverb *gerade* 'right-now' attributively modify the theme argument of transitive verbs and in such case these are interpreted with the meaning of current state as is shown in (55b).

(55) a. ? der von der Mutter gerade geküsste Junge
the by the.Dat mother.F right-now kiss.PastPart baby
'the baby being kissed by the mother' (German)

b. das vom Mädchen gerade gelesene Buch
the by girl.F right-now read.PastPart book
'the book being read by the girl' (German)

c. ? die gerade dem Mädchen gegebene Blume
the right-now the.Dat girl.F give.PastPart flower
'the flower being given to the girl' (German)

(56) a. علی نے کچھ پرندے اڑتے (ہوئے) دیکھے.
ɑli=ne kʊc^h pərɪnde ʊɽ-te (hu-e) dek^h-e
Ali.M=Erg some bird.M.3Pl fly-Imperf.M.3Pl (be-Perf.M.3Pl) see-Perf.M.3Sg
'Ali saw some birds flying.'

b. ندا نے مجرم کو پٹتے (ہوئے) دیکھا.

nıda=ne	mʊjım=ko	pıṭ-te	(hu-e)	dek^{h}-a
Nida.F=Erg	criminal=Acc	beaten-Imperf.Obl	(be-Perf.Obl)	see-Perf.M.3Sg

‘Nida saw the criminal being beaten.’

c. علی نے ایک سیب گرتا (ہوا) دیکھا.

ɑli=ne	ek	seb	gır-ta	(hu-a)	dek^{h}-a
Ali.M=Erg	one	apple.M.3Sg	fall-Imperf.M.3Sg	(be-Perf.M.3Sg)	see-Perf.M.3Sg

‘Ali saw an apple falling.’

d. ندا نے ماں کو بچے کو چومتے (ہوئے) دیکھا.

nıda=ne	mã=ko	bace=ko	cum-te	(hu-e)	dek^{h}-a
Nida.F=Erg	mother=Acc	baby=Acc	kiss-Imperf.Obl	(be-Perf.Obl)	see-Perf.M

‘Nida saw the mother kissing the baby.’

An adverbial use of characterizing present participles is shown in (56). Here we see that if the object, which the present participle characterizes, is unmarked, the participle agrees with the object in number and gender (56a,c). And if the object is marked, the oblique form of the participle is used (56b,d). Characterizing present participles are not used predicatively, in contrast to adjectives.

4.7.2 Resultative participles

Nedjalkov & Jaxontov (1988), after having undertaken a crosslinguistic study of resultative constructions,[6] defined the resultatives as ‘those verb forms that express a state implying a previous event’. A resultative participle characterizes its head ‘by expressing a state that results from a previous event’ (Haspelmath 1994:159). The resultative participles are used attributively and predicatively crosslinguistically. The latter construction is sometimes referred to as *be*-passives. It has been argued (Frajzyngier 1978:154) that there is no difference between *be*-passives and general copula sentences crosslinguistically, except that the predicate in *be*-passives is morphologically derived from the lexical class of verbs.

[6]The term resultative refers to two constructions i.e. resultative participles and resultative secondary predicates (Embick 2004:360). Here I am talking about the first one.

Unlike English and many other languages, resultative and passive participles in Urdu are formed in the syntax via auxiliaries. The resultative participle is formed from the perfect participle of the main verb followed by the perfect participle of the verb *ho* ‘be/become’. In a passive participle, on the other hand, the perfect participle of the verb *ja* follows the perfect participle of the main verb.

(57) a1. آئے (ہوئے) لوگ
a-e (hu-e) log
come-Perf.M (be-Perf.M) people
‘People who have come’

a2. Die angekommenen Leute
the come.PastPart people.Pl
‘People who have come’

b1. پٹا (ہوا) مجرم
pıṭ-a (hu-a) mʊjrım
be-beaten-Perf (be-Perf) criminal
‘The criminal, who has been beaten’

b2. No equivalent intransitive verb

c1. گرا (ہوا) درخت
gır-a (hu-a) dɑrɑxt
fall-Perf.M (be-Perf.M) tree.M
‘A fallen tree’

c2. Der gefallene Baum
the fall.PastPart tree.M
‘The fallen tree’

(58) a1. بچے کو ڈسا (ہوا) سانپ
*bɑcce=ko dɑs-a (hu-a) sãp
baby=Acc bite-Perf (be-Perf) snake
‘The snake, who has bitten the baby’

a2. No agent oriented participle

b1. سانپ کا ڈسا (ہوا) بچہ
sãp=ka dɑs-a (hu-a) bɑccɑ
snake=Gen.M bite-Perf.M (be-Perf.M) baby.M.Sg
‘The baby bitten by a snake’

b2. Das von der Schlange gebissene Baby
the.Neut by the.Dat.F snake.F.Sg bite.PastPart baby.Neut.Sg
‘The baby bitten by a snake’

c1. کتاب پڑھی (ہوئی) لڑکی
kıtab pɑṛ^h-i (hu-i) lɑṛki
book read-Perf.F (be-Perf.F) girl.F
‘The girl, who has read the book’

c2. No agent oriented participle

d1. لڑکی کی پڑھی (ہوئی) کتاب
lɑṛki=ki pɑṛ^h-i (hu-i) kıtab
girl.F read-Perf.F (be-Perf.F) book
‘The book read by the girl’

d2. Das vom Mädchen gelesene Buch
The by girl read.PastPart book
‘The book read by the girl’

Resultative participles of some intransitive verbs from Urdu and German are illustrated in (57) and of some transitive verbs are illustrated in (58). The resultative participles in Urdu are typical in two respects. For one, they can be oriented either towards the agent or the patient/theme. Secondly, while modifying the patient/theme they can optionally take the agent as the genitive marked argument.

In (58a1), the resultative participle takes accusative object and cannot have active orientation. Haspelmath (1994) with reference to Subbarao (1984) has reported that the resultative participles taking only bare NP complements in Urdu have active orientation as shown in (59).

(59) دس لڈّو کھایا ہوا لڑکا
dɑs lɑḍḍu **k^ha-ya** **hu-a** lɑṛka
ten laddu.M.3Sg eat-Perf.M.3Sg be-Perf.M.3Sg boy.M.3Sg
'A boy who ate ten laddus (sweet meats)' (Subbarao 1984:170)

	Intransitive Verb			Transitive Verb		
Orientation	Agent (unerg)	Patient	Theme (unacc)	Agent	Patient	Theme
Urdu	yes	yes	yes	yes	yes	yes
German	yes	N.A	yes	no	yes	yes

Table 4.6: Attributive use of resultative participles

A summary of attributive use of resultative participles in Urdu and German is given in Table 4.6. In the case of ditransitive verbs, the resultative participles only orient to the theme argument, as shown in (60a-b). The genitive marked agent in participles of ditransitive verbs is obligatorily realized. In Urdu dative subject verbs also have theme oriented resultative participles, as shown in (60c).

(60) a. علی کا دیا ہوا انعام
ɑli=ka dɪ-ya (hu-a) ɪn'am
Ali.M=Gen give-Perf.M (be-Perf.M) prize.M
'The prize given by Ali'

b. علی کا ندا کو دیا ہوا انعام

ɑli=ka nɪda=ko dɪ-ya (hu-a) ɪn'am
Ali.M=Gen Nida.F=Dat give-Perf.M (be-Perf.M) prize.M
'The prize given by Ali to Nida'

c. ندا کو ملا ہوا انعام

nɪda=ko mɪl-a (hu-a) ɪn'am
Nida.F=Dat attain-Perf.M (be-Perf.M) prize.M
'The prize attained by Nida'

(61) a. ندا کی مرتّبہ کتاب

nɪda=ki **mʊrattabah** kɪtab
Nida.F.3Sg=Gen.F composed book.F.3Sg
'The book composed by Nida' (Arabic-loaned participle)

b. ندا کی مرتّب کردہ کتاب

nɪda=ki **mʊrattab-kardah** kɪtab
Nida.F.3Sg=Gen.F composed book.F.3Sg
'The book composed by Nida' (Persian-loaned participle)

c. ندا کی مرتّب کی ہوئی کتاب

nɪda=ki **mʊrattab** **k-i** **hu-i** kɪtab
Nida.F.3Sg=Gen.F compose-Perf.F.3Sg do-Perf.F be-Perf.F.3Sg book.F.3Sg
'The book composed by Nida' (Urdu resultative participle)

In addition to native Urdu participles, some participles from Arabic and Persian are also used as adjectives in Urdu. The use of three participle forms from Arabic, Persian and Urdu is illustrated in (61a-c).

There is one difference between resultative participial forms used as adjectives in Urdu and other argument-taking adjectives taken from Arabic and Persian in that the former ones being participial forms of certain transitive verbs can take bare NPs as their complements and the latter ones, like English adjectives, do not allow bare NPs in their argument structure at all.

The use of resultative participles in clausal constructions is illustrated in (62). All instances with the unmarked subject seem to resemble the subject-complement construction of a copula sentence. Each of these constructions has a stative reading that has resulted after the action is executed/performed. The adverbial use of resultative participles is illustrated in (63). Only with some matrix verbs, resultative participles can be used adverbially.

(62) a. بہت لوگ آئے ہوئے ہیں۔

bɑhʊt	log	a-e	hu-e	hɛ̃
many	people	come-Perf.3Pl	be-Perf	be.3Pl.Pres

‘Many people are in the state of having come.’

b. ایک درخت گرا ہوا ہے۔

ek	dɑrɑxt	gɪr-a	hu-a	hɛ
one	tree	fall-Perf.M	be-Perf.M	be.3Sg.Pres

‘One tree is fallen’

c. ایک لڑکی (نے) کتاب پڑھی ہوئی ہے۔

ek	lɑɽki=(ne)	kɪtab	pɑɽʰ-i	hu-i	hɛ
one	girl.F=(Erg)	book.F.3Sg	read-Perf.F.3Sg	be-Perf.F	be.3Sg.Pres

‘One girl has read the book’

d. یہ کتاب اس لڑکی کی پڑھی ہوئی ہے۔

yɪh	kɪtab	ʊs	lɑɽki=ki	pɑɽʰ-i	hu-i	hɛ
this	book.F.3Sg	that	girl.F.3Sg=Gen.F	read-Perf.F.3Sg	be-Perf.F	be.3Sg.Pres

‘This book is read by that girl’

(63) علی کو کھلونا ٹوٹا ہوا ملا۔

ɑli=ko	kʰɪlona	ʈuʈ-a	hu-a	mɪl-a
Ali.M=Dat	toy.M.3Sg	break-Perf.M.Sg	be-Perf.M.Sg	meet-Perf.M.3Sg

‘Ali found the toy broken.’

In (64) the participle forms are of intransitive verbs and the genitive marked elements are not subject arguments. These genitive marked elements give adverbial meanings and must be classified as adjuncts.

(64) a. رات کا آیا ہوا

rat=ka night.F.3Sg=Gen.M.3Sg | a-ya come-Perf.M.3Sg | hu-a be-Perf.M.3Sg

'One who has come at night'

b. دودھ کا جلا ہوا

dud^{h}=ka milk.M.3Sg=Gen.M.3Sg | jɑl-a burn-Perf.M.3Sg | hu-a be-Perf.M.3Sg

'One, who has been burnt by milk'

c. لاڈ کا پلا ہوا

laḍ=ka love.M.3Sg=Gen.M.3Sg | pɑl-a be-brought-up-Perf.M.3Sg | hu-a be.Perf.M.3Sg

'One who has been brought up with love'

Some participle forms of intransitive verbs with incorporated argument are used as adjectives as in *sɑr p^{h}ɪra* 'mad' (lit: head revolved), *nak kaṭa* 'one whose nose has been cut'.

4.8 Summary

The verb *ho* 'be/become' in Urdu encodes many notions or meanings in different structures. In this chapter different uses of *ho* were analyzed: as a copula, as a dynamic verb, as tense auxiliary, as a light verb and its uses in subordinate clauses and participial adjuncts. There are some syntactic tests that distinguish stative *ho* from dynamic *ho*. Syntactic frames of stative copula and dynamic copula were explored. Many subcategorization frames of stative copula were identified. The dynamic copula only takes a subset of syntactic frames of stative copula. It was shown that postural stative verbs and dynamic verbs in Urdu can be distinguished by their different distribution in subordinate clauses involving the use of *ho*. The characterizing participles were discussed in view of arguments they modify and the arguments they take in their argument structure.

Chapter 5

Arguments and syntax of nouns

Like verbs, some nouns usually derived from verbs also take arguments, however, their number and case-marking might not be the same as that of the corresponding verbs. This chapter explores arguments of nouns in Urdu. In section 1, different types of case phrases a noun can take as its arguments are described. The genitive marked arguments/adjuncts of nouns are discussed in detail in section 2. Multiple instances of genitive marked elements in noun phrases are analysed and it is shown that some nouns in Urdu can take two genitive marked arguments. Discontinuous constituents result in Urdu NPs when an argument-taking noun is modified by an argument-taking adjective. Argument taking adjectives in Urdu are first discussed in section 3 and then the phenomenon of NP-internal discontinuity in Urdu is analyzed and implemented in LFG in section 4. Section 5 concludes the chapter.

5.1 Argument-taking nouns

Chomsky (1970) first showed that verbs and nouns seem to share complement-taking properties (1).

(1) a. The enemy destroyed the city.

b. The enemy's destruction of the city.

One basic difference between the argument structure of verbs and nouns is that verbs can take bare NPs as their arguments, but nouns cannot.

Argument taking nouns in Urdu are mostly derived from verbs of Arabic origin, some are of Persian origin and a very few are of Urdu origin. Different case marked nouns as well as prepositional phrases can be arguments of nouns in Urdu. A list of such elements with examples of nouns taking them as arguments is given in Table 5.1.

Nr.	Type of Argument	Example of Noun Phrase
(i)	Dative Marked	hɑkʊmɑt=ko **ɪntɪbah** government.F.3Sg=Dat warning.F.3Sg 'warning to the government'
(ii)	Instrumental Marked	bɛ'ɑt=se **ɪnkar** obedience.F.3Sg=Inst denial 'denial of obedience'
(iii)	Locative (in) Marked	abadi=mẽ **ɪzafɑh** population.F.3Sg=Loc.in increase.M.3Sg 'increase in population'
(iv)	Locative (on) Marked	fɔj=pɑr **ɪnhɪsar** army.F.3Sg=Loc.on dependence.M.3Sg 'dependence on army'
(v)	Locative (till) Marked	cɑ̃d=tɑk **rɑsai** moon.M.3Sg=Loc.till reach.F.3Sg 'reach up to the moon'
(vi)	Adpositional	sɑlamti bare **tɑʃviʃ** security.F.3Sg about anxiety.F.3Sg 'anxiety over security'
(vii)	Genitive Marked	tavan=ka **mʊtalbɑh** ransom.M.3Sg=Gen.M.3Sg demand.M.3Sg 'demand for ransom'

Table 5.1: Argument-taking nouns

All the bold-faced nouns in Table 5.1 are in fact taken from Arabic, except the noun *rasai* 'reach' which is derived from a Persian verb *rasidan* 'to reach'. These nouns form constituents with their arguments and appear in one structural position in a sentential clause. For example, the noun plus its arguments can take the subject position in copular constructions (2). The head noun and its to-be-illustrated argument are bold-faced in (2). *ko* marked, *se* marked, *par* marked, *mẽ* marked, *tak* marked, adpositional (*bare*) and genitive marked arguments are illustrated in (2a) to (2g), respectively. Although sometimes there are other arguemtns/adjuncts of the head noun, only one is bold-faced for illustration.

(2) a. دہشت گردوں کے خلاف کاروائی پر پاکستان کا امریکا کو انتباہ مضحکہ خیز ہے.

[dɑhʃɑt-gɑrdõ=ke xılaf karvai=pɑr pakıstan=ka
[terrorist.3Pl=Gen.Obl against action.F.3Sg=Loc.on Pakistan=Gen.M.3Sg
ɑmrika=ko ıntıbah] mɑzhɑkɑh-xez hɛ
America=Dat warning.M.3Sg] ridiculing be.Pres.3Sg
'Pakistan's warning to America on action against terrorists is ridiculing.'

b. امام حسین کا یزید کی بیعت سے انکار فطری تھا.

[ımam hʊsɛn=ka [**yɑzid=ki bɛ'ɑt**]=**se**
[Imam Hussain=Gen.M.3Sg [**Yazid=Gen.F.3Sg obedience.F.3Sg**]=**Abl**
ınkar fıtri t^ha
denial.M.3Sg] natural be.Past.Sg
'Imam Husain's denial to obey Yazid was natural.'

c. جوہری توانائی پر انحصار صحیح نہیں ہے.

[[**jɔhri tɑvanai**]=**pɑr ınhısar**] sɑhih nɑhĩ hɛ
[[**atomic energy.F.3Sg**]=**Loc.on dependence.M**] right not be.Pres.3Sg
'Dependence on atomic energy is not right.'

d. آبادی میں اضافہ بڑا مسئلہ ہے.

[**abadi=mẽ ızafɑh**] bɑṛa mɑs'lɑh hɛ
[**population.F.3Sg**=Loc.in **increase.M**] big problem.M.3Sg be.Pres.3Sg
'Increase in population is a big problem.'

e. سورج تک رسائی ممکن نہیں.

[**sʊrɑj=tɑk rɑsai**] mʊmkın nɑhĩ
[**sun.M.3Sg=Loc.till reach.F.3Sg**] possible not
'Approach to the sun is not possible.'

f. عوام کی ملکی سلامتی بارے تشویش بلاجواز نہیں.

[ɑvam=ki [**mʊlki sɑlamti**] **bare tɑʃviʃ**] bıla-javaz nɑhĩ
[people=Gen.F [**of-country security.F**] **about anxiety.F**] without-reason not
'People's anxiety over country's security is not without reason.'

g. سرائیکی قوم کا صوبہ سرائیکستان کا مطالبہ جائز ہے۔

[[sɑraiki	qɔm]=ka	**[subɑh**	**sɑraikɪstan]=ka**	
[[Saraiki	nation.F.3Sg]=Gen.M	**[province.M.3Sg**	**Saraikistan]=Gen.M.3Sg**	
mʊtalbɑh]	jaɪz	hɛ		
demand.M.3Sg]	rightful	be.Pres.3Sg		

'Saraiki nation's demand for Saraikistan province is rightful.'

Some examples of native nouns which take arguments are *b*ʰ*arosah* 'trust', *baṛ*ʰ*otri* 'increase', *pahõc* 'reach' and *mãg* 'demand', etc. The last three nouns are from verbs *baṛ*ʰ 'increase', *pahõc* 'reach' and *mãg* 'ask' respectively. Some borrowed nouns from English also take arguments in Urdu, for example, the noun *barifĩg* takes the *par* 'on' marked argument in Urdu.

5.2 Genitive modifiers/arguments

In this section the syntactic structure of those noun phrases of Urdu is explored in which there are multiple instances of genitive marked elements. It will be shown that different genitive elements in noun phrases result into the different hierarchical structure. Due to morpho-syntactic properties of the genitive marker and agreement requirements different structural ambiguities are generated. Before discussing the genitive arguments, issues with the genitive modifiers are introduced. It is shown that the attributive genitive modifiers stack together at the same level to modify the head noun (Raza, 2010c). The nominals which license genitive marked arguments are then described and classified.

5.2.1 The genitive marker

The genitive marker in Urdu-Hindi originated from the past participle form of the Indo-Aryan verb *kar-* 'do'; it inflects for gender, number and case and agrees with the head noun: *ka* (M.Sg.Dir), *ke* (M.Sg.Obl/M.Pl), *ki* (F) (Payne, 2004). The form *ka* is traced back to a Prakrit past participle form *kera* 'done' that is further traced back to a Sanskrit past participle form *krita*. The evidence for it is provided by Beames (1879) who provides the example, *kapi-kritam vacanam* 'speech made by monkey' or alternatively 'speech of the monkey'. The inflected forms *kera, keri, kere* were in use in Old Hindi. Possessive/genitive forms of pronouns were made by adding these forms to them. Later, the first syllable of these forms was lost and only the second one was retained with some forms of the pronouns and hence Hindi-Urdu now uses *mera, tera*, etc. as possessive pronouns. But with some other forms of pronouns and all nouns the forms *ka, ki, ke* began to be used. The most frequent use of these forms is that they

mark possessive nouns, that is, these generally express possession or a *have-a* relation. Consider the following instances of genitive phrases.

(3) a. ندا کی کتاب
nıda=ki kıtab
Nida.F.3Sg=Gen.F book.F.3Sg
'Nida's book'

b. کتاب کا ورق
kıtab=ka vɑrq
book.F.3Sg=Gen.M.3Sg page.M.3Sg
'The page of a book'

Both instances in (3) show a *have-a* relation. Although the genitive markers are hosted by the modifier noun, they show agreement of number and gender with the head noun. Based on some tests to distinguish affixes from clitics by Miller (1992) and Zwicky (1987), Butt & King (2005) have analyzed these forms as clitics. One of the tests is that these have scope over noun coordination (4a) and the other is that some other element can intervene between these endings and the nominal host (52b).

(4) a. ندا کی کتاب اور کاپی
nıda=ki kıtab ɔr kapi
Nida.F.3Sg=Gen.F book.F.3Sg and note-book
'Nida's book and note-book'

b. ندا ہی کی کتاب
nıda=hi=ki kıtab
Nida.F.3Sg=Emp=Gen.F book.F.3Sg
'Nida's book'

In addition to agreement, another requirement for a genitive phrase to be grammatical is that the host of the genitive marker should be in the oblique form.

(5) a. ندا کا بچّہ

nɪda=ka	bɑcca
Nida.F.3Sg.Obl=Gen.M.3Sg	child.M.3Sg.Dir

‘Nida’s child’

b. بچّے کا کھلونا

bɑcce=ka	kʰɪlona
child.M.3Sg.Obl=Gen.M.3Sg	toy.M.3Sg.Dir

‘The toy of the child’

c. کھلونے کی قیمت

kʰɪlone=ki	qimɑt
toy.M.3Sg.Obl=Gen.F	price.F.3Sg

‘The price of the toy’

(6) a. ندا کے بچّے کا کھلونا

[nɪda=ke(*ka)	bɑcce(*bɑcca)]=ka	kʰɪlona
[Nida.F.3Sg.Obl=Gen.M.3Sg.Obl	child.M.3Sg.Obl]=Gen.M.3Sg.Dir	toy.M.3Sg

‘The toy of Nida’s child’

b. بچّے کے کھلونے کی قیمت

[bɑcce(*bɑcca)=ke(*ka)	kʰɪlone(*kʰɪlona)]=ki	qimɑt
[child.M.3Sg.Obl=Gen.M.3Sg.Obl	toy.M.3Sg.Obl]=Gen.F	price.F.3Sg

‘The price of the child’s toy’

In all of instances in (5) the hosts of the genitive markers are in the oblique form. Proper nouns and feminine singular common nouns usually have morphologically identical direct and oblique forms. The host in (5a) is the proper noun and here it is supposed that it is in oblique form. The host of the genitive marker *bacca* ‘child’ in (5b) and *kʰɪlona* ‘toy’ are in their oblique forms, respectively, as *bacce* and *kʰɪlone*. If a genitive phrase hosts a genitive marker then both the head noun and the modifier alongside the genitive marker in the host genitive phrase are in oblique form (6). The head noun *kʰɪlona* ‘toy’ in (6a) is modified by the phrase *nɪda=ka bɑcca* ‘Nida’s child’

which must be realized as *nɪda=ke bacce*. Similarly the head noun *qimat* 'price' in (6b) is modified by the phrase *bacce=ka kʰɪlona* 'child's toy' which must be realized as *bacce=ke kʰɪlone*.

In addition to the possession relation there are many other relations that can be expressed by genitive markers in Urdu (Platts, 1967), for example, kinship relation, part-whole relationship, etc.

5.2.2 Structure of noun phrases with multiple genitive modifiers

Both flat and hierarchical structures of genitive phrases with multiple instances of genitive marked nouns are possible. Consider the following example.

(7) a. علی کی چاندی کی انگوٹھی

ɑli=ki	cɑ̃di=ki	ɑ̃guṭʰi
Ali.M.3Sg=Gen.F	silver.F.3Sg=Gen.F	ring.F.3Sg

'Ali's silver-ring' OR
'The ring of Ali's silver'

b. ندا کا سونے کا کنگن

nɪda=ka	sone=ka	kɑ̃gɑn
Nida.F.3Sg=Gen.M	gold.M.3Sg=Gen.M.3Sg	bracelet.M.3Sg

'Nida's golden ring'

For (7a), the following three bracketing structures could all be assumed.

(i) [ɑli=ki [cɑ̃di=ki ɑ̃guṭʰi]]
(ii) [[ɑli=ki cɑ̃di]=ki ɑ̃guṭʰi]
(iii) [ɑli=ki cɑ̃di=ki ɑ̃guṭʰi]

The first two are the more plausible structures for a genitive phrase and are both hierarchical. In (i) the head noun *ɑ̃guṭʰi* 'ring' is modified by the genitive marked element *cɑ̃di=ki* 'of silver' and then the resulting genitive phrase is modified by another genitive marked element *ɑli=ki* 'of Ali'. In the second bracketing structure (ii), first

the noun *cɑ̃di* 'silver' is modified by *ɑli=ki* 'of Ali' and then the genitive marker is attached to this phrase to modify the head noun *ɑ̃guṭʰi* 'ring'. In (iii) the head noun *ɑ̃guṭʰi* 'ring' is modified by two genitives *ɑli=ki* 'of Ali' and *cɑ̃di=ki* 'of silver'.

Three structures for a noun phrase with two genitive modifiers are not always possible. For example, in (7b) the bracketing structure [[*nɪda=ka sone*]=*ka kɑ̃gan*] is not possible due to morpho-syntactic behavior of the genitive markers described in the previous section. Here, a genitive marker is supposed to be attached with another genitive phrase *nɪda=ka sone* 'Nida's gold' in which the genitive marker *ka* does not show the oblique agreement of case with the head noun *sone* 'gold'. So, a hierarchical structure like (ii) is completely out for (7b).

The flat structure for both instances in (7) is actually not plausible in that the two genitive modifiers cannot alternate their positions as in (8). So, we cannot say that the two modifiers are modifying the head noun at the same level.

(8) a. چاندی کی علی کی انگوٹھی

* cɑ̃di=ki	ɑli=ki	ɑ̃guṭʰi
silver.F.3Sg=Gen.F	Ali.M.3Sg=Gen.F	ring.F.3Sg

'Ali's silver-ring'

b. سونے کا ندا کا کنگن

* sone=ka	nɪda=ka	kɑ̃gɑn
gold.M.3Sg=Gen.M.3Sg	Nida.F.3Sg=Gen.M	bracelet.M.3Sg

'Nida's golden ring'

One could argue that (8) is ungrammatical due to another reason, namely that the possessive modifier *ɑli=ki* 'of Ali' is not the most prominent (the outer most). However, in (9) there is no possessive modifier and still only one order (9a) of genitive marked elements is grammatical. Some more instances of genitive phrases where a flat structure of genitive modifiers is not possible are given in (10).

(9) a. ملتان کے مٹّی کے برتن

mʊltan=ke	mɑṭṭi=ke	bɑrtɑn
Multan=Gen.M.Pl	clay=Gen.M.pl	pot.M.Pl

'Ceramic pots of Multan'

b. مٹّی کے ملتان کے برتن

* mɑʈʈi=ke mʊltan=ke bɑrtɑn
clay=Gen.M.pl Multan=Gen.M.Pl pot.M.Pl
'Ceramic pots of Multan'

(10) a. علی کی بیوی کی انگوٹھی

ɑli=ki bivi=ki ãguʈʰi
Ali=Gen.F wife.F.3Sg=Gen.F ring.F.3Sg
'The ring of Ali's wife'

b. علی کی گھڑی کی قیمت

ɑli=ki gʰɑɽi=ki qimɑt
Ali=Gen.F watch.F.3Sg=Gen.F price.F.3Sg
'The price of Ali's watch'

The noun that opens a position for another nominal is called the relational element (Seiler, 1983). The noun *bivi* 'wife' in (10a) is a relational element and the modifier *ali=ki* makes a constituent with this noun. In (10b) the head noun *qimat* 'price' actually is an attribute and this attribute can only be of *gʰaɽi* 'watch'. So, (10a) and (10b) both have a deep hierarchical structure on the left (due to our world knowledge), as illustrated below.

(i) [[ɑli=ki bivi]=ki ãguʈʰi]
(ii) [[ɑli=ki gʰɑɽi]=ki qimɑt]

The hierarchical structure could be deep on either side depending upon the agreement of case and the semantics of the participants in genitive phrases. Some examples of genitive phrases where the structure is deep on the right side are given in (11) below.

(11) a. اونٹ کی دوڑنے کی رفتار

ũʈ=ki [dɔɽ-ne=ki rɑftar]
camel.M.3Sg=Gen.F run-Inf.Obl=Gen.F speed.F.3Sg
'The running speed of a camel'

b. ندا کی سکول کی گاڑی

nɪda=ki	[sɑkul=ki	gaṛi]
Nida.F.3Sg=Gen.F	school.M.3Sg=Gen.F	bus.F.3Sg

'Nida's school bus'

(12) a. اونٹ کے دوڑنے کی رفتار

[ũṭ=ke	dɔṛ-ne]=ki	rɑftar
camel.M.3Sg=Gen.M.Obl	run-Inf.Obl=Gen.F	speed.F.3Sg

'The speed of running of the camel'

b. ندا کے سکول کی گاڑی

[nɪda=ke	sɑkul]=ki	gaṛi
Nida.F.3Sg=Gen.M.Obl	school.M.3Sg=Gen.F	bus.F.3Sg

'The bus of Nida's school'

(13) a. اونٹ کے دوڑنے کے انداز

ũṭ=ke	dɔṛ-ne=ke	ɑndaz
camel.M.3Sg=Gen.M.3Sg	run-Inf.Obl=Gen.M.3Sg	style.M.3Pl

'The running styles of a camel' OR 'The styles of running of a camel'

b. ندا کی یونیورسٹی کی گاڑی

nɪda=ki	university=ki	gaṛi
Nida.F.3Sg=Gen.F	university.F.3Sg=Gen.F	bus.F.3Sg

'Nida's university-bus' OR 'the bus of university of Nida'

The morpho-syntactic properties of the genitive marker, that is, the oblique form of the host and the agreement rule out the left-sided deep structure in (11). Likewise the right-sided deep structure is out in (12). However, in (13) the structure can be deep on either side because the morpho-syntactic properties and agreement are ambiguous.

It was shown here that different forms of the genitive marker agreeing with the head noun in such constructions entail different meanings. Due to more than one case of agreement ambiguity is generated.

5.2.3 Attributive genitive modifiers

A list of some attributes and some of their possible types are given in Table 5.2. Attributive genitive modifiers here are taken to be those genitive modifiers which result after adding a genitive marker to some attribute of the head noun. More than one attributive genitive element can modify the head noun at the same level. That is, a genitive phrase with multiple attributive genitive modifiers has a flat structure.

Nr.	Attribute	Type
1	material	gold, clay
2	price	10 rupees, low price
3	size	small size
4	height	tall height
5	color	red color
6	weight	low wight, 80 gram
7	age	20 years, less age

Table 5.2: List of attributes

Consider the following instances of genitive phrases with multiple instances of attributive genitive modifiers.

(14) a. لمبے قد کی گورے رنگ کی لڑکی

[lɑmbe qad=ki] [gore rɑ̃g=ki] lɑṛki
[tall.M.3Sg height.M.3Sg=Gen.F] [white.M.3Sg color.M.3Sg=Gen.F] girl.F.3Sg
'A girl of tall height and white color'

b. کم قیمت کا چھوٹے سائز کا لیپ ٹاپ

[kɑm qimɑt=ka] [cʰoṭe saɪz=ka] laptop
[low price.F.3Sg=Gen.M.3Sg] [small size.M.3Sg=Gen.M.3Sg] laptop.M.3Sg
'A laptop of low price and small size'

(15) a. گورے رنگ کی لمبے قد کی لڑکی

[gore rɑ̃g=ki] [lɑmbe qad=ki] lɑṛki
[white.M.3Sg color.M.3Sg=Gen.F] [tall.M.3Sg height.M.3Sg=Gen.F] girl.F.3Sg
'A girl of white color and tall height'

b. چھوٹے سائز کا کم قیمت کا لیپ ٹاپ

[c^{h}oṭe saɪz=ka] [kɑm qimɑt=ka] laptop
[small size.M.3Sg=Gen.M.3Sg] [low price.F.3Sg=Gen.M.3Sg] laptop.M.3Sg
'A laptop of small size and low price'

In both examples of (14a) and (14b) the order of genitive modifiers can be changed without altering the truth-conditional meaning. The examples with the changed order are shown in (15). The claim, here, is that attributive genitive modifiers stack together in a flat structure in Urdu. Furthermore, attributive genitive modifiers show a syntactic distribution similar to adjectival modifiers. The meanings conveyed by attributive genitives in (14) and (15) can also be conveyed by adjectives, as shown in (16) and (17).

(16) a. لمبی گوری لڑکی

lɑmbi gori lɑṛki
tall.F.3Sg white.F.3Sg girl.F.3Sg
'A tall and white girl'

b. سستا چھوٹا لیپ ٹاپ

sɑsta c^{h}oṭa laptop
cheap.M.3Sg small.M.3Sg laptop.M.3Sg
'A cheaper and smaller laptop'

(17) a. گوری لمبی لڑکی

gori lɑmbi lɑṛki
white.F.3Sg tall.F.3Sg girl.F.3Sg
'A white and tall girl'

b. چھوٹا سستا لیپ ٹاپ
c^{h}oṭa sɑsta laptop
small.M.3Sg cheap.M.3Sg laptop.M.3Sg
'A smaller and cheaper laptop'

Another syntactic distribution that is shared by both adjectival modifiers and attributive genitive modifiers is that they must appear adjacent to the head noun they are modifying. Any genitive modifiers that also appear in the NP constituent must be ordered before the adjectives. This is illustrated with the adjectival modifiers *baṛa* 'big' and *saf* 'clean' in (18)–(19) below.

(18) a. علی کا بڑا کمرہ
ɑli=ka bɑṛa kɑmrɑ
Ali.M.3Sg=Gen.M.3Sg big.M.3Sg room.M.3Sg
'Ali's big room'

b. بڑا علی کا کمرہ
* bɑṛa ɑli=ka kɑmrɑ
big.M.3Sg Ali.M.3Sg=Gen.M.3Sg room.M.3Sg
'Ali's big room'

(19) a. پینے کا صاف پانی
pine=ka saf pani
drink.Inf.Obl=Gen.M.3Sg clean water.M.3Sg
'Purified drinking-water'

b. صاف پینے کا پانی
saf pine=ka pani
clean drink.Inf.Obl=Gen.M.3Sg water.M.3Sg
'Purified drinking-water'

The head noun *kamra* 'room' in (18a) is modified by the adjective *baṛa* 'big' and then the resulting phrase is modified by *ali=ka* 'of Ali' and the whole phrase is grammatical. However, in (18b) the adjectival modifier *baṛa* 'big' is not modifying the head

noun *kamra* 'room'. We cannot suppose that this adjective is modifying the possessor noun because in that case it should have been in oblique form with 'Ali', which is oblique given that it is hosting a genitive *ka*. As the adjective cannot modify the genitive phrase *ali=ka kamra* 'Ali's room', the whole phrase becomes ungrammatical. Similarly, the phrase in (19b) can only be grammatical if the phrase *pine=ka pani* 'drinking-water' is considered as a unit. As in English, the phrase every men's room' is acceptable because 'men's room' is considered to be a noun-noun compound involving the possessive morpheme and 'every' takes 'men's room' as a unit for its complement (Baker 1995:6-7). Likewise the noun phrases in (8) and (9b) are also ungrammatical because the attributive genitive modifiers are not modifying the head nouns in these. These examples are repeated in (20) for convenience.

(20) a. چاندی کی علی کی انگوٹھی

* cãdi=ki	ɑli=ki	ãguṭʰi
silver.F.3Sg=Gen.F	Ali.M.3Sg=Gen.F	ring.F.3Sg

'Ali's silver-ring'

b. سونے کا ندا کا کنگن

* sone=ka	nɪda=ka	kãgɑn
gold.M.3Sg=Gen.M.3Sg	Nida.F.3Sg=Gen.M	bracelet.M.3Sg

'Nida's golden ring'

c. مٹّی کے ملتان کے برتن

* mɑṭṭi=ke	mʊltan=ke	bɑrtɑn
clay=Gen.M.pl	Multan=Gen.M.Pl	pot.M.Pl

'Cermaic pots of Multan'

Sometimes ambiguity is generated as to whether the genitive attributive modifier before the material genitive modifier is for the material of the head noun or the head noun itself (21a). If the material genitive modifier is placed before any other genitive attributive modifier (21b) then no such ambiguity is generated.

(21) a. سرخ رنگ کی لکڑی کی میز

sʊrx	rɑ̃g=ki	lɑkṛi=ki	mez
red	color.M.3Sg=Gen.F	wood.F=Gen.F	table.F.3Sg

‘The table made of red wood’ OR
‘The red wooden table’

b. لکڑی کی سرخ رنگ کی میز

lɑkṛi=ki	sʊrx	rɑ̃g=ki	mez
wood.F=Gen.F	red	color.M.3Sg=Gen.F	table.F.3Sg

‘The red wooden table’

When both adjectives and attributive genitive modifiers are present in noun phrases then attributive genitive modifiers are placed near the head noun before the adjectives, as shown in (22).

(22) a. ندا کا سفید رنگ کا خوبصورت لباس

nɪda=ka	sʊfɛd	rɑ̃g=ka	xubsurɑt	lɪbas
Nida.F=Gen.M.3Sg	white	color.M.3Sg=Gen.M.3Sg	beautiful	suit.M.3Sg

‘Nida’s beautiful suit of white color’

b. ندا کے ملتان کے مٹّی کے رنگین برتن

nɪda=ke	mʊltan=ke	mɑṭṭi=ke	rɑ̃gin	bɑrtɑn
Nida.F=Gen.M.Pl	Multan=Gen.M.Pl	clay.F=Gen.M.Pl	colorful	pot.M.Pl

‘Nida’s colorful ceramic pots of Multan’

The order of different genitive modifiers and adjectives in Urdu noun phrases can be generalized as the following:

possessor + relational genitives + attributive genitives + adjectives + noun

There are some more crucial points with respect to the attributive genitives. For one, the part-whole relation which is usually expressed by marking the whole with the genitive marker, can also be expressed by attributing some specified part to the whole by marking the specified part for genitive case. In the first case the genitive phrase is

either the possessor or the whole (23) and in the latter case it should be considered an attributive genitive (24).

(23) a. مرغے کی ٹانگ
mʊrɣe=ki ṭãg
rooster.M.3Sg=Gen.F leg.F.3Sg
'The leg of a rooster'

b. پنکھے کے دو پر
pãkʰe=ke do pɑr
fan.M.3Sg=Gen.M.Pl two wing.M
'Two wings of a fan'

(24) a. ایک ٹانگ کا مرغا
ek ṭãg=ka mʊrɣa
one leg.F.3Sg=Gen.M.3Sg rooster.M.3Sg
'The rooster of one leg'

b. پیلے رنگ کا دو پروں کا یہ پنکھا
pile rãg=ka do pɑrõ=ka yɪh pãkʰa
yelloe color.M=Gen.M two wing.M.3Pl=Gen.M this fan.M.3Sg
'This fan of yellow color and of two wings'

Another point is that some attributive genitives contain either a specified attribute or its concrete value to modify countable nouns, as shown in (25)–(26). However, such attributive genitives cannot modify uncountable nouns, as shown in (27).

(25) a. کم قیمت کا لیپ ٹاپ
kɑm qimɑt=ka laptop
low price.F.3Sg=Gen.M.3Sg laptopt.M.3Sg
'A laptop of low price'

b. پندرہ سو روپے کا لیپ ٹاپ

pãdrɑh sɔ rupɛ=ka laptop
fifteen hundred rupee.M.3Pl =Gen.M.3Sg laptop.M.3Sg
'A laptop of fifteen hundred rupees'

(26) a. کم وزن کی روٹی

kɑm vɑzɑn=ki roṭi
low weight.M.Sg=Gen.F bread.F.Sg
'A bread of low weight'

b. اسی گرام کی روٹی

ɑsi gɑram=ki roṭi
eighty gram=Gen.F bread.F.Sg
'A bread of 80 grams'

(27) a. کم (وزن کی) چینی

kɑm (*vɑzɑn=ki) cini
less (*weight.M.Sg=Gen.F) sugar.F
'Less sugar'

b. اسی گرام (کی) چینی

ɑsi gɑram=(*ki) cini
eighty gram=(*Gen.F) sugar.F
'80 gram sugar'

Finally, abstract nouns cannot have attributive genitive modifiers. In (28a) and (28b) the modified nouns are abstract and therefore modifiers in such cases are not attributive genitive modifiers. An *is-a relation* is instead expressed in (28b).

(28) a. قیمت کا علم

qimɑt=ka ılm
price.F.3Sg=Gen.M.3Sg knowledge.M.3Sg
'Knowledge of the price'

b. سو روپے کی قیمت

sɔ	rupɛ=ki		qimɑt
hundred	rupee.M.3Pl	=Gen.F.3Sg	price.F.3Sg

'Price of one hundred rupees'

So far, in this section, multiple instances of genitive modifiers have been explored and the analogy of attributive genitive modifiers with adjectives has been analyzed. It has been shown that the attributive genitive modifiers stack at the same level otherwise there is an hierarchical structure when there are some genitives encoding inherent relations and possession. An order of different modifiers in noun phrases was established. Genitive arguments of nouns are discussed in the next section.

5.2.4 Nominals and genitive arguments

Some nouns like brother, child, enemy, edge, etc., are supposed to be inherently relational (Partee & Borschev, 2003), taking genitive arguments in English and many other languages, but the clearest cases of arguments in noun phrases are found in nominalizations. First it was believed that nouns take arguments only optionally (Anderson, 1983; Dowty, 1987). Later it was shown by Grimshaw (1990) that many nouns have two senses or interpretations. In one sense they denote complex events and take arguments obligatorily and in another sense they denote simple events and do not necessarily take arguments. In the first sense they are called process nominals or derived nominals (Babby, 1997) and in the latter sense they are called result nominals. Later in this section, it will be examined whether this distinction also exists in Urdu.

5.2.4.1 Infinitives with genitive arguments

In Urdu, infinitives are used to construct clauses and are also used as nominals. Butt (1995) has debated whether infinitive clauses in Urdu are nominalizations or gerunds. Bhatt (2005) proposed that infinitives can be projected without a subject but Davison (2008) recommended only a full clause structure for Urdu infinitives. She provided evidence for the presence of a projected syntactic subject in infinitive clauses, even if it is not pronounced. I would say infinitives in Urdu can act both as nouns and as clauses.

In this section, Urdu infinitives as nominals will be explored as to what types of genitive arguments they can take. Consider the example phrases of infinitival nominals in (29)–(31).

(29) a. ایندھن کا جلنا

ĩdʰɑn=ka jɑl-na
fuel.M.3Sg=Gen.M.3Sg burn-Inf.M.Sg
'Burning of fuel' (Unaccusative)

b. مریض کا کھانسنا

mɑriz=ka kʰãs-na
patient=Gen.M.3Sg cough-Inf.M.Sg
'Coughing of the patient' (Unergative)

c. مجرم کا پٹنا

mʊjrɪm=ka pɪṭ-na
criminal=Gen.M.3Sg be-beaten-Inf.M.Sg
'Being beaten of the criminal'

(30) a. تصاویر کا دیکھنا

tɑsavir=ka dekʰ-na
picture.F.Pl=Gen.M.3Sg see-Inf.M.Sg
'Seeing of pictures'

b. ندا کا دیکھنا

nɪda=ka dekʰ-na
Nida=Gen.M.3Sg see-Inf.M.Sg
'Seeing of Nida/ Seeing by Nida'

c. علی کا ندا کو دیکھنا

ɑli=ka nɪda=ko dekʰ-na
Ali=Gen.M.3Sg Nida=Acc see-Inf.M.Sg
'Seeing of Nida by Ali'

For both unaccusative intransitive verbs (29a) and unergative intransitive verbs (29b), the subject argument (the theme in the first case and the agent in the latter case) of the nominal is marked genitive. Likewise the patient argument of the intransitive

verb *pɪṭ* 'be-beaten' is marked for genitive case. The question is which argument of the transitive verb *dekʰ* 'see' is marked by genitive marker in (30b). Is it subject or object? It is assumed that it can be either in Urdu. If only one argument of the infinitival nominal of a transitive verb is mentioned in Urdu then it can be either its internal argument or its external argument.

Lebeaux (1986), however, notes that if the subject of a nominal derived from a transitive verb is present, then the object must obligatorily be there for the noun phrase to be grammatical. Grimshaw showed that obligatory arguments are taken by nominals only when these are action nominals and it could also be the case that the same nominal behaves in both senses. The nominal in (30b) will therefore be considered a result nominal. We observe that with Urdu infinitives only one argument is marked genitive. In case both the subject and the object arguments of a transitive verb are realized, the subject is marked genitive and the object is marked nominative or accusative. This is illustrated in (31).

(31) a. ندا کا بیچنا

nɪda=ka bec-na
Nida=Gen.M.3Sg sell-Inf.M.Sg
'Selling by Nida/ Selling of Nida'

b. کھلونوں کا بیچنا

kʰɪlonõ=ka bec-na
toy.M.3Pl.Obl=Gen.M.3Sg sell-Inf.M.Sg
'Selling of toys'

c. ندا کا کھلونے بیچنا

nɪda=ka kʰɪlone bec-na
Nida=Gen.M.3Sg toy.3Pl sell-Inf.M.Sg
'Selling of toys by Nida'

d. ندا کا بچّوں کو کھلونے بیچنا
nɪda=ka baccõ=ko kʰɪlone bec-na
Nida=Gen.M.3Sg child.3Pl toy.M.3Pl sell-Inf.M.Sg
'Selling of toys by Nida to children'

(32) a. مالدار کا دینا
maldar=ka de-na
wealthy=Gen.M.3Sg give-Inf.M.Sg
'Giving by the wealthy'

b. فقیر کو دینا
fɑqir=ko de-na
beggar.M=Dat give-Inf.M.Sg
'Giving to a beggar'

c. خیرات (کا) دینا
xɛrat=(ka) de-na
alms.F=Gen.M give-Inf.M
'Giving of alms'

d. فقیر کو خیرات کا دینا
fɑqir=ko xɛrat=(ka) de-na
beggar.M=Dat alms.F=Gen.M give-Inf.M.Sg
'Giving alms to a beggar'

e. مالدار کا فقیر کو دینا
maldar=ka fɑqir=ko de-na
wealthy=Gen.M.3Sg beggar.M=Dat give-Inf.M.Sg
'Giving to a beggar by the wealthy'

f. مالدار کا خیرات دینا
maldar=ka xɛrat de-na
wealthy=Gen.M.3Sg alms.F give-Inf.M
'Giving of alms by the wealthy'

g. مالدار کا فقیر کو خیرات دینا

maldar=ka fɑqir=ko xɛrat de-na
wealthy=Gen.M.3Sg beggar.M=Dat alms.F give-Inf.M.Sg
'Giving alms to a beggar by the wealthy'

In (31), the infinitive of the transitive verb *bec* 'sell' is given with its arguments. The subject reading of genitive modifier in (31a) and object reading in (31b) both are okay. In (31c–d) both object and subject are included and only the subject is marked genitive. The indirect object of a ditransitive verb like *de* 'give' is never marked for genitive case in the infinitival construction, as shown in (32). If the subject argument is not realized, the direct object can be unmarked or marked for genitive case, else it is obligatorily unmarked.

(33) a. ایک خدا کا ماننا

ek xʊda=ka man-na
one God=Gen.M.3Sg believe-Inf.M.Sg
'Believing in one God'

b. قران کا پڑھنا

qʊran=ka pɑɽʰ-na
Quran.M.3Sg=Gen.M.3Sg read-Inf.M.Sg
'Reading of the Quran'

The instances in (33) are frequent expressions in Urdu web corpora where only the object reading is construed. In case of only a single genitive argument of the infinitive, most of the times, the object reading is meaningful.

We have seen that either the subject or the object argument in Urdu infinitives is marked for genitive case, but not both simultaneously. In this respect Urdu is similar to Persian, where infinitives can take either subject or object, but not both with the *ezafe* construction (Ghomeshi, 1997). In Urdu, however, not only are some arguments of infinitival nominals marked for genitive case, but arguments of verbs in participial adjectives and in some subordinate clauses are also marked genitive. Participial adjectives have already been discussed in Chapter 4 under the heading of characterizing participles.

5.2.4.2 Other nominals with genitive marked arguments

In Urdu many nouns other than infinitves are derived from verbal roots and take genitive arguments. Some nouns are derived from verbal roots of Urdu itself and some are derived from verbal roots of other languages like Arabic and Persian. These nouns can be divided into two classes. The nouns in one class take only one genitive marked argument and the nouns in the other class can take two genitive marked arguments. Some instances of nouns from the first class are given in (34)–(35).

(34) a. ٹرین کی روانگی
train=ki rɑvangi
train.F.3Sg=Gen.F.3Sg departure.F.3Sg
'Departure of the train'

b. ٹرین کی اسٹیشن سے روانگی
train=ki station=se rɑvangi
train.F.3Sg=Gen.F.3Sg station.M.3Sg=Abl departure.F.3Sg
'Departure of the train from the station'

(35) a. سیلاب کی تباہی
sɛlab=ki tɑbahi
flood.M.3Sg=Gen.F.3Sg destruction.F.3Sg
'Destruction due to floods'

b. فصلوں کی تباہی
fɑslõ=ki tɑbahi
crop.F.3Pl=Gen.F.3Sg destruction.F.3Sg
'Destruction of crops'

c. انسان کی تباہی
ɪnsan=ki tɑbahi
man.M.3Sg=Gen.F.3Sg destruction.F.3Sg
'Destruction of man / destruction by man'

d. انسان کی فصلوں کی تباہی

* ɪnsan=ki fɑslõ=ki tɑbahi
man.M.3Sg=Gen.F.3Sg crop.F.3Pl=Gen.F.3Sg destruction.F.3Sg
'Destruction of crops by man'[1]

e. سیلاب کی فصلوں کی تباہی

* sɛlab=ki fɑslõ=ki tɑbahi
flood.M.3Sg=Gen.F.3Sg crop.F.3Pl=Gen.F.3Sg destruction.F.3Sg
'Destruction of crops due to floods'

f. انسان کے ہاتھوں فصلوں کی تباہی

ɪnsan=ke hatʰõ fɑslõ=ki tɑbahi
man.M.3Sg=Gen.Obl hand.Pl.Obl crop.F.Pl destruction.F.3Sg
'Destruction of crops by man'

g. سیلاب سے فصلوں کی تباہی

tsɛlab=se fɑslõ=ki tɑbahi
flood.M.3Sg=Abl crop.F.3Pl=Gen.F.3Sg destruction.F.3Sg
'Destruction of crops due to floods'

In (34) *ravangi* 'departure' is a noun derived form the Persian word *ravanɪh* 'dispatched' and has two alternate subcategorization frames. In both cases it takes a genitive marked subject. All nouns of an intransitive nature have their subject as genitive marked. The noun *tabahi* 'destruction' is of a transitive nature and can take either a genitive marked subject or a genitive marked object (35a–c) but both cannot be marked for genitive case simultaneously (35d–e). In (35f) the agent argument is encoded by adpositional phrase and in (35g) the instrumental marked noun encodes the cause of the event. Other nominals in Urdu which fall in this class are, for example, *pɪtai* 'beating', *dʰʊlai* 'washing', *pɪsai* 'crushing', *mʊ'aɪnah* 'examination'.

There are some nouns that can take only a genitive marked object, for example, the nominal *bacao* 'safety' derived from the verb *baca* 'save'. The noun *ɪntɪxab* 'selection' usually takes a genitive object. However, it can also act as a result nominal where it refers to the result of the selection process.

[1] This phrase is grammatical in another reading: 'Destruction of crops of man'. In such case the bracketing structure would be as: [*ɪnsan=ki faslõ*]=*ki tabahi*

(36) a. صدر کا انتخاب
sɑdɑr=ka ɪntɪxab
president=Gen.M.3Sg selection.M.3Sg
'Selection of the president'

b. علی کا انتخاب
ɑli=ka ɪntɪxab
Ali.M.3Sg=Gen.M.3Sg selection.M.3Sg
'Selection made by Ali'

c. یہ خوبصورت شعر علی کا انتخاب ہے
yɪh xubsurɑt ʃer ɑli=ka ɪntɪxab hɛ
this beautiful verse Ali=Gen.M.3Sg selection.M.3Sg be.Pres
'This beautiful verse is a selection made by Ali'

In (36b), the noun *ɪntɪxab* 'selection' refers to some result of the process which is evidenced in (36c). Because event or process nominals cannot be used predicatively, as showed by Grimshaw (1990), the instance of the noun *ɪntɪxab* 'selection' in (36c) is a result nominal which is modified by a genitive modifier.

In addition to the above nouns, there is another class of nominals in Urdu which take two genitive marked arguments. The noun *g^herao* 'surrounding' derived from the verb *g^her* 'surround' is one example. For each of the head nouns in (37), there are two arguments and both are marked genitive. Urdu infinitives can be marked for case like other nouns. In (37b), the second argument is infinitival.

(37) a. نوجوانوں کا تھانے کا گھیراؤ
nɔjɑvanõ=ka t^hane=ka g^herao
youngster.M.3Pl=Gen.M.3Sg police-station=Gen.M.3Sg surround.M.3Sg
'Surrounding of police-station by youngsters'

b. صدر کا الیکشن کرانے کا اعلان
sɑdɑr=ka [election kɑr-a-ne]=ka ɪ'lan
president=Gen.M [election.M.3Sg do-Caus-Inf.Obl]=Gen.M announcement.M
'Announcement made by the president to conduct elections'

c. عوام کا صدر کے اعلان کا خیر مقدم

ɑvam=ka	[sɑdɑr=ke	ɪ'lan]=ka	xɛr-mɑqdɑm
people=Gen.M	president=Gen.M	[announcement.M]=Gen.M	welcome.M.Sg

'Welcome by the people of the announcement of the president'

It is typical for the second class of nominals in Urdu in that both subject and object/theme are marked by genitive markers. In Persian, there is not a single nominal in which both subject and object are licensed by the *ezafe* construction. In English too, both subject and object of a noun cannot be realized prenominal genitives simultaneously. It is observed in Urdu and some other Indo-Aryan languages (for example, Saraiki) that both subject and object/theme in some nominals are marked for genitive case. Also in German, deverbal nouns of a transitive nature allow for two genitives one in the prenominal position and the other in the postnominal position (Butt *et al.*, 1996).

5.2.5 Implementation of NPs with multiple genitives in LFG

In an LFG grammar of Urdu (Butt & King, 2007) genitive markers are dealt with within syntax. Genitive clitics have their own terminal node to represent the head of their case phrase. The agreement of the genitive marker in gender, number and case with the head noun is dealt with at f-structure via feature unification. The genitive phrases with a single instance of genitive marked element work well. With multiple instances of genitive elements the complexity increases. Consider (38):

(38) a. ابرہا کا ہاتھیوں کا لشکر

ɑbrɑha=ka	hatʰɪyõ=ka	lɑʃkɑr
Abraha.M.3Sg=Gen.M.3Sg	elephant.M.3Pl=Gen.M.3Sg	army.M.3Sg

'Abraha's army of elephants'

b. ابرہا کے ہاتھیوں کا لشکر

ɑbrɑha=ke	hatʰɪyõ=ka	lɑʃkɑr
Abraha.M.3Sg=Gen.M.3Pl	elephant.M.3Pl=Gen.M.3Sg	army.M.3Sg

'The army of Abraha's elephants'

c. ابرہا کے ہاتھیوں کے لشکر
ɑbrɑha=ke hatʰɪyõ=ke lɑʃkɑr
Abraha.M.3Sg=Gen.M.3Pl elephant.M.3Pl=Gen.M.3Pl army.M.3Pl
'Abraha's armies of elephants' OR 'The armies of Abraha's elephants'

Consider the following three bracketed structures. For (38a), the bracketed structure in (i) is valid; for (38b), the bracketed structure in (ii) is valid and for (38c) two bracketed structures in (i) and (ii) are valid.

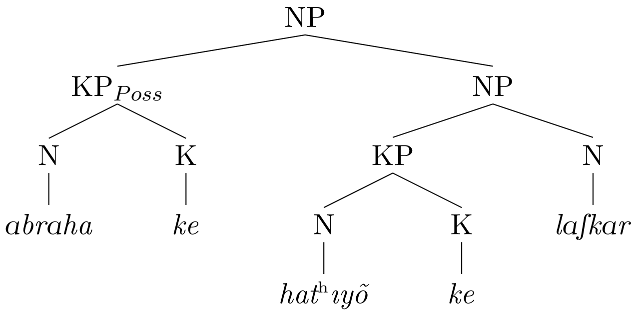
(i) [ɑbrɑha=kX [hatʰɪyõ=kX lɑʃkɑr]]
(ii) [[ɑbrɑha=kX hatʰɪyõ]=kX lɑʃkɑr]
(iii) [ɑbrɑha=kX hatʰɪyõ=kX lɑʃkɑr]

If no restrictions of agreement and/or hierarchy are posited at f-structure, all three bracketed structures for any instance in (38) should result. Having put the agreement restrictions, the bracketed structure in (ii) will be out for (38a) via feature unification, as the first case form *ka* in it does not agree with the noun *hatʰɪyõ* 'elephants'. Now the issue is how to generate the hierarchical structure for it as in (i) rather than the flat structure as in (iii). A hierarchical structure in genitive phrases is usually found when there are possessors or inherently relational nouns. I would suggest encoding features of nouns in the grammar, e.g., whether some noun is inherently relational and whether something is able to be possessed. All nouns cannot be possessed, for example, the noun 'price' cannot be possessed. If some noun is inanimate then it cannot be the possessor. The noun 'Abraha' is a human name, so it qualifies as a possessor. Furthermore, features for recognizing attributive genitives should also be devised. Attributive genitives always stack together at the same level. So, working with these features the flat structure for (38a) would be out and only the hierarchical structure in (i) would result.

The agreement morphology will also rule out the bracketed structures in (i) and (iii) for (38b). The second case form *ka* in it is singular which agrees with the singular noun *lɑʃkɑr* 'army'. The first genitive form *ke* should then definitely be in agreement with the plural noun *hatʰɪyõ* 'elephants'. Hence only the bracketed structure (ii) would result for (38b) via feature unification.

Example (38c) is crucial in that two hierarchical structures are possible for it because the agreement of case markers is ambiguous. The first case marker *ke* agrees with the noun *hatʰɪyõ* 'elephants' and also with the plural noun *lɑʃkɑr* 'army'. It should be noted that the singular direct and the plural direct forms of many masculine common nouns in Urdu are identical. The plural form of the noun *lɑʃkɑr* 'army' in this example is

recognized due to the second case form *ke*. The valid c-structures and the corresponding f-structure representations of (38c) are shown in Figures 5.1 and 5.2[2]. The right-sided deep structure has the corresponding f-structure with SPEC and ADJUNCT and the left-sided deep structure has the corresponding f-structure only with ADJUNCT.

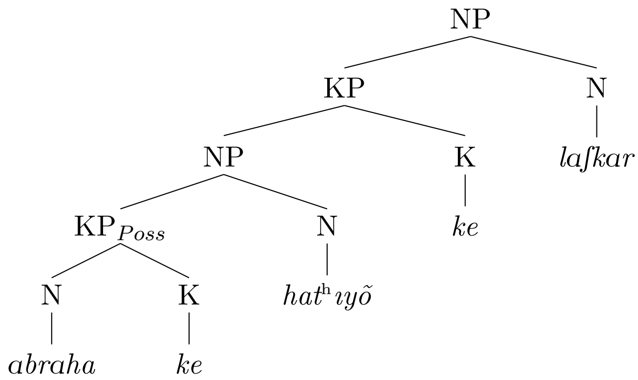


PRED	'lɑʃkɑr'	
SPEC	PRED	'ɑbrɑha'
	CASE	gen
	NUM	sg
ADJUNCT	{ PRED	'hat^{h}i'
	CASE	gen
	GEND	masc, NUM pl, PERS 3 }
GEND masc,	NUM pl, PERS 3	

Figure 5.1: A valid c- and f-structure representation of (38c)

Another issue is that a non-ambiguous noun phrase with multiple genitives may become ambiguous when it itself is marked by a case. This is due to the fact that direct forms of nouns and cases in a noun phrase are changed to oblique forms when a case clitic is added. For example, when the accusative case marker *ko* is added to to any instance of (38a–b), both result in an identical surface structure as in (38c) and the resulting phrase has two meanings (39).

[2]I use the simplified f-structures in my discussions.



PRED	‘laʃkar’		
ADJUNCT	{ PRED	‘hatʰi’	
	SPEC	PRED	‘abraha’
		CASE	gen
		NUM	sg
	CASE	gen	
	GEND	masc, NUM pl, PERS 3 }	
GEND masc,	NUM pl, PERS 3		

Figure 5.2: Another valid c- and f-structure representation of (38c)

(39) ابرہا کے ہاتھیوں کے لشکر کو روکو!

[ɑbrɑha=ke	hatʰıyõ=ke	laʃkɑr]=ko	rok-o
Abraha.M.3Sg=Gen.M.3Pl	elephant.M.3Pl=Gen.M.3Pl	army.M.3Pl	stop-Imper

‘Stop Abraha’s army of elephants!’ OR
‘Stop the army of Abraha’s elephants!’

In Table 5.3 different types of genitive elements in a noun phrase with the proposed equivalent functional categories are given. The pseudo-possessives are as the first genitive in *nıda=ki sakul=ki gaṛi* ‘Nida’s school bus’ and inherently relational genitives are as in *nıda=ka bʰai* ‘Nida’s brother’.

As the attributive genitive modifiers syntactically show the same distribution as simple adjectives do, these should be treated as adjuncts in the f-structure. Consider another genitive phrase in (40) with different types of genitive elements. The f-structure of this phrase is shown in Figure 5.3. A feature ‘adjunct-type’ can be added to show the particular kind of genitive adjunct.

Nr.	**Type of genitive**	**Functional category**
1	Genitive$_{possessive}$	SPEC
2	Genitive$_{pseudo-possessive}$	SPEC
3	Genitive$_{inherently-relational}$	SPEC
4	Genitive$_{relational}$	ADJUNCT
5	Genitive$_{attributive}$	ADJUNCT

Table 5.3: Different genitive elements with functional labels

(40) ندا کے ملتان کے نیلے رنگ کے مٹّی کے قیمتی برتن

nɪda=ke	mʊltan=ke	nile	rɑ̃g=ke	mɑṭṭi=ke
Nida.F=Gen.M.Pl	Multan=Gen.M.Pl	blue	color=Gen.M.Pl	clay.F=Gen.M.Pl

qimɑti	bɑrtɑn
expensive	pot.M.Pl

'Nida's expensive blue ceramic pots from Multan'

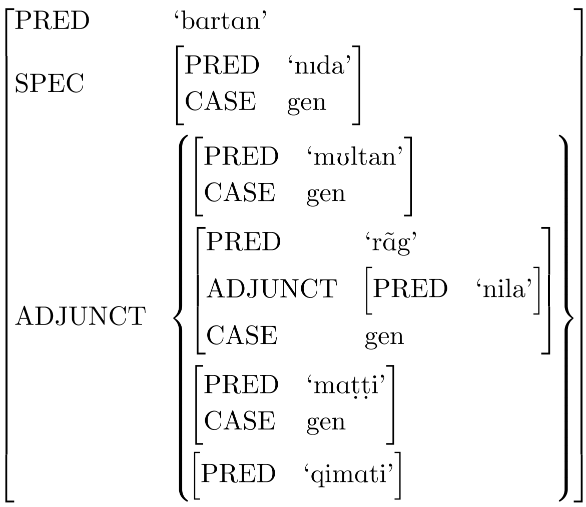


Figure 5.3: f-structure representation of (40)

For deverbal nouns it is proposed that they should be listed in the lexicon according to the number and type of genitive arguments they take. The lexicon with full subcategorization information of nouns will help to correctly parse the noun phrases in Urdu and the coverage of the parser will be increased.

5.2.6 Summary

In this section, nominals in Urdu with genitive adjuncts and genitive arguments have been analyzed. Different classes of nominals based on the number and type of genitive arguments have been investigated. The flat and hierarchical structure of such phrases was explored. It was shown that attributive genitive modifiers behave like adjective modifiers in the syntax and can stack together at the same level. It was proposed to deal with such elements separately from other genitive elements like possessive genitives and inherently relational genitives. Some nouns take either genitive marked agent or genitive marked theme/object, others take agent and theme/object both marked for genitive case.

Different types of genitives in Urdu noun phrases and different possible hierarchical structures of such phrases is an interesting phenomenon to be investigated further in syntactic theories formally.

5.3 Argument-taking adjectives

In Urdu, participial adjectives and some other adjectives originally derived from verbal stems of other languages also take arguments. Participle forms of verbs used as adjectives in Urdu were discussed in detail in section 4.7 under the heading of characterizing participles.

Some of the argument-taking adjectives are listed in Table 5.4. The nouns in parenthesis are modified by the argument taking adjectives. Nouns are put in parenthesis because the purpose here is to illustrate only the adjectives and their arguments. These adjectives usually are derived from Arabic verbal stems and can sometimes be replaced with native participial adjectives as shown in (41).

(41) a. صدر کو حاصل اختیارات

sɑdɑr=ko	hasıl	ıxtiyarat
president=Dat	attained	power.M.3Pl

'Powers attained by the president'

Nr.	Type of Argument	Example of Adjective Phrase
(i)	Dative Marked	sɑdɑr=ko **hasıl** (ıxtıyarat) president=Dat attained (powers) '(The powers) attained by the president'
(ii)	Instrumental Marked	ɑdlıyɑh=se **xaıf** (hʊkmɑran) courts=Inst afraid (rulers) '(The rulers) afraid of courts'
(iii)	Locative (in) Marked	bʊxar=mẽ **mʊbtɑla** (ʃɑxs) fever=Loc.in suffered (man) '(The man) suffered from fever'
(iv)	Locative (on) Marked	tɑqarir=pɑr **mʊʃtɑmıl** (kıtab) speech.Pl=Loc.on comprised (book) '(The book) comprised of speeches'
(v)	Adpositional	sıhɑt=ke lıye **mʊzır** (xurak) health=Gen for harmful (food) '(The food) harmful for health'
(vi)	Genitive Marked	sɑdɑr=ke **hami** (ɑfrad) president=Gen.Pl supporting (people) '(People) supporting the president'

Table 5.4: Argument-taking adjectives

b. صدر کو ملے ہوئے اختیارات

sɑdɑr=ko mıl-e hu-e ıxtiyarat
president=Dat attain-Perf.M.3Pl be-Perf.M.3Pl power.M.3Pl
'Powers attained by the president'

The adjective *hasıl* in (41a) and the participial form *mıl-e hu-e* in (41b) are semantically equivalent with the meaning of 'attained'.[3]

(42) a. صدر کی حامی جماعت نے ریلی نکالی.

sɑdɑr=ki hami jɑma'ɑt=ne reli nıkal-i
president=Gen.F supporting party.F.3Sg=Erg rally.F.3Sg conduct-Perf.F.3Sg
'The supporting party of the president conducted a rally'

[3]The equivalent construct in Persian for this meaning is *hasıl ʃʊdah* which is the participial form of *hasıl ʃʊdan*. In Urdu, the adjective *hasıl* can also be considered as the reduced participle form of complex predicate *hasıl ho*.

b. صدر کے حامی افراد نے ریلی نکالی.
sɑdɑr=ke hami ɑfrad=ne reli nɪkal-i
president=Gen.F supporting people.3Pl=Erg rally.F.3Sg conduct-Perf.F.3Sg
'The supporting people of the president conducted a rally'

c. صدر کے حامیوں نے ریلی نکالی.
sɑdɑr=ke hamɪyõ=ne reli nɪkal-i
president=Gen.F supporter.3Pl=Erg rally.F.3Sg conduct-Perf.F.3Sg
'Supporters of the president conducted a rally'

In English -*er* nouns like 'supporter' are not used as adjectives but their equivalents in Urdu are also used as adjectives as shown in (vi) of Table 5.4. They are used both as attributive adjectives (42a–b) and as nouns (42c). When these are used as nouns, these can be inflected like nouns. They are not inflected when used as adjectives. The genitive case clitics that mark the arguments of such elements are inflected according to the number and gender of the (modified) noun (42a–c). In addition to adjectives similar to -*er* English nouns, some other participial forms of verbs from Arabic, Persian and many from Urdu also take a genitive marked subject argument and are used as adjectives.

Traditionally it was believed that adjectives do not subcategorize for an OBJ function crosslinguistically. Later it was found that adjectives can take accusative complements in some languages (first observed by Platzack (1982) for old Swedish). Such adjectives were called transitive adjectives (Maling, 1983). In Urdu, apparently there are no adjectives that take bare NPs as their complements. However, participle forms of transitive verbs can take bare NPs to modify the subject argument of the verb. More cases of complement taking adjectives are described below.

5.3.1 Genitive marked complements of degree adjectives

The theme of some degree adjectives in Urdu is also marked by the genitive case, as is shown in (43) for the two adjectives *bʊra* 'bad' and *acc*ʰ*a* 'good'.

(43) a. دل کا برا شخص

dɪl=ka	bʊra	ʃɑxs
heart.M.3Sg=Gen.M.3Sg	bad.M.3Sg	person.M.3Sg

‘a person of bad heart’

b. طبیعت کی اچھی لڑکی

tɑbi’ɑt=ki	ɑcc^{h}i	lɑɽki
nature.F.3Sg=Gen.F.3Sg	good.F.3Sg	girl.F.3Sg

‘a girl of good nature’

Ikeya (1995) argued that degree adjectives in English are one-place predicates and that the contextual/semantic dimensions are in fact modifiers of these predicates. He reported three contextual dimensions of such predicates: Thematic Dimension (TD), Comparative Dimension (CD) and Degree Dimension (DD). In his example *He is very good at basketball for a short Japanese* all these dimensions are expressed: *at basketball* is TD, *for a short Japanese* is CD and *very* is DD. The first dimension TD was first reported by Bartsch (1986/87). The comparative and thematic dimensions of degree adjectives in Urdu are usually encoded by different case phrases or adpositional phrases. The genitive marked elements in (43) in fact are modifiers of adjectives encoding the thematic dimension of adjectives.

Constructions similar to (43) also exist in Welsh (Mittendorf & Sadler, 2008) and Arabic (Al-Sharifi & Sadler, 2009). Morris-Jones (1931) has called genitives in such construction as ‘genitives of respect’ or ‘in respect of genitives’. Mittendorf & Sadler (2008) have analysed the genitive of respect in attributive constructions as the OBJ function of the adjective. On similar lines the the f-structure representation of (43b) is given in Figure 5.4.

$$\left[\begin{array}{ll} \text{PRED} & \text{‘lɑɽki’} \\ \text{ADJUNCT} & \left\{ \left[\begin{array}{ll} \text{PRED} & \text{‘ɑcc}^{h}\text{i’} \\ \text{OBJ} & \left[\begin{array}{ll} \text{tɑbi’ɑt} & \\ \text{CASE} & \text{gen} \end{array}\right] \end{array}\right] \right\} \end{array}\right]$$

Figure 5.4: f-structure representation of (43b)

5.3.2 Dative marked complements of degree adjectives

Consider some copular sentences with dative elements of Urdu in (44) and of German in (45).

(44) a. علی کو یہ قمیض ڈھیلی ہے۔
ɑli=ko yɪh qɑmiz d^{h}ili hɛ
Ali.M=Dat this shirt.F.3Sg loose.F.3Sg be.3Sg.Pres
'This shirt is loose to Ali.'

b. ندا کو وہ پاجامہ تنگ ہے۔
nɪda=ko vʊh pajamɑh tɑ̃g hɛ
Nida.F=Dat that trouser.M.3Sg tight be.3Sg.Pres
'That trouser is tight to Nida.'

(45) a. Dieses Hemd ist ihm zu weit
this.Neut shirt be.3Sg.Pres Pron.3Sg.M.Dat too loose
'This shirt is too loose to him.'

b. Diese Hose ist ihr zu eng
this.F trouser be.3Sg.Pres Pron.3Sg.F.Dat too tight
'This trouser is tight to her.' (German)

The dative NPs in (44) can be dropped and still the instances remain grammatical. Those datives in German and in some other languages that may be dropped without any syntactic and semantic residue are called free datives (e.g. Hole 2009). Hole has given the following test to distinguish free datives from sub-categorized datives in German:

(46) *A dative argument* D not dependent on a preposition is *free* in a simple positive declarative sentence S of German **iff**

(i) S without D is grammatical;

(ii) S without D does not entail that there is an individual

(α) which participates in the event described by S and

(β) which could be encoded as a dative argument. (Hole 2009:3)

The Urdu datives in (44) also act like the free datives of German. As the free datives in these construction encode the beneficiaries, in the f-structure representation these can be analyzed as OBJ-GO of adjectives as is shown in Figure 5.5.

It should be noted that copular verb in (44) cannot be assumed to introduce a genitive marked argument in addition to PREDLINK because in that case it would not be called a copular verb. The dative argument is due to the adjectival predicate, hence it should be encoded as an argument of the adjective.

$$
\begin{bmatrix}
\text{PRED} & \text{`ho}\left\langle (\uparrow \text{SUBJ})\ (\uparrow \text{PREDLINK}) \right\rangle\text{'} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`pajamɑh'} \\ \text{SPEC} & \text{`vʊh'} \\ \text{PERS} & \text{3, NUM sg} \end{bmatrix} \\
\text{PREDLINK} & \begin{bmatrix} \text{PRED} & \text{`tɑ̃g'} \\ \text{OBJ-GO} & \left\{ \begin{bmatrix} \text{PRED} & \text{`nɪda'} \\ \text{PERS} & \text{3, CASE dat} \end{bmatrix} \right\} \end{bmatrix} \\
\text{TENSE} & \text{present}
\end{bmatrix}
$$

Figure 5.5: f-sructure representation of (44b)

5.3.3 Clausal complements

Adjectives in Urdu can also allow for clausal complements. There is one class of adjectives that can appear in the copula constructions illustrated in (47). The morphemes *yɪh* or *yɪh bat* are equivalent to expletives in Urdu and can sometimes be dropped.

(i) [Nominalized property] Adj Cop

(ii) [(*yɪh/yɪh bat*) 'it'] Adj Cop CP

(47) a. علی کا انعام جیتنا ممکن ہے.

[ɑli=ka ɪn'am jit-na] **mʊmkɪn** hɛ
[Ali=Gen.M.3Sg prize.M.3Sg win-Inf.M] possible be.Pres.3Sg
'Ali's winning of the prize is possible.'

b. (یہ (بات)) ممکن ہے کہ علی انعام جیتے.

(yɪh (bat)) **mʊmkɪn** hɛ [kɪh ɑli ɪn'am jit-e]
(this (thing.F.3Sg)) possible be.Pres [that Ali prize.3Sg win-Subjn]
'It is possible that Ali will win the prize.'

In addition to *mʊmkɪn* 'possible', some other adjectives that fall in this class are *yaqini* 'sure', *zaruri* 'important', *sahih* 'true', etc.

Consider another class of adjectives which can take three alternate frames with the typical example *hɛran* 'surprised' illustrated in (48).

(i) NP-*par*

(ii) that-clause

(iii) NP-*par* that-clause

(48) a. ندا علی کے انعام جیتنے پر حیران ہے۔

nɪda [ɑli=ke ɪn'am jit-ne]=pɑr **hɛran** hɛ
Nida.F [Ali=Gen.Obl prize.M win-Inf.Obl]=Loc.on surprised be.Pres.3Sg
'Nida is surprised at Ali's winning of the prize.'

b. ندا حیران ہے کہ علی نے انعام جیتا۔

nɪda **hɛran** hɛ [kɪh ɑli=ne ɪn'am jit-a]
Nida.F surprised be.Pres.Sg [that Ali=Erg prize.M win-Perf.M]
'Nida is surprised that Ali won the prize.'

c. ندا اس بات پر حیران ہے کہ علی نے انعام جیتا۔

nɪda ɪs bat=pɑr **hɛran** hɛ [kɪh ɑli=ne ɪn'am jit-a]
Nida this thing=Loc.on surprised be.Pres.Sg [that Ali=Erg prize.M win-Perf.M]
'Nida is surprised that Ali won the prize.'

d. ندا علی پر حیران ہے کہ اس نے انعام جیتا۔

nɪda ɑli=pɑr **hɛran** hɛ [kɪh ʊs=ne ɪn'am jit-a]
Nida.F Ali.M=Loc.on surprised be.Pres.Sg [that 3Sg=Erg prize.M win-Perf.M]
'Nida is surprised that Ali won the prize.'

It can be said that the canonical argument of the adjective *hɛran* 'surprised' is *par* marked NP. However, if the *par* marked element is some nominalization then the adjective can alternatively take *that*-clause arguments (48b–d). The adjectives *pareʃan* 'sad' and *xʊʃ* 'happy' show similar syntactic behavior.

5.4 Discontinuous constituents within noun phrases

Clause-level discontinuity of noun phrases in different languages has been reported in the literature under different headings like discontinuity, extraposition, extraction, free topic and quantifier float. However, discontinuity within the constituent-level has not been discussed before. Urdu is a language in which both types of discontinuity, clause-level and constituent-level exist. The arguments of heads in NPs are non-contiguous to their respective heads at surface structure. This detachment can occur at one structural position of a clause inside the NP rather than spanning the whole clause with other constituents in between. The heads in such phrases cannot precede their arguments. This discontinuity and the constraints on the order of elements in noun phrases in Urdu pose a modeling problem. Syntactic analysis of such noun phrases is made. In contrast to what is usually assumed within ParGram (Butt *et al.*, 1999; Dipper, 2003), a flat c-structure for Urdu NPs is therefore proposed and is modeled in the LFG framework.

5.4.1 State of the Art

In simple terms, a constituent in which the individual words that make it up appear separately and interleaved with other elements is called a discontinuous constituent. Languages that exhibit several of the following characteristics: free word order, discontinuous constituents, split-ergative case marking and null anaphora, are often referred to as non-configurational languages in the syntactic literature (Hale, 1980, 1983, 1992; Simpson, 1983, 1991). Warlpiri is considered to be a typical example of a non-configurational language in Australian aboriginal languages and shows almost all the properties that have been associated with non-configurational languages. A sample sentence of Warlpiri is given in (49) to show the phenomenon of discontinuous noun phrases across the clause.

(49)	**wita-jarra-rlu**	ka-pala	wajili-pi-nyi	yalumpu	**kurdu-jarra-rlu**
	small-Du-Erg	Pres-3DuSubj	chase-NPast	that.Abs	child-Du-Erg
	maliki.				
	dog.Abs				

‘The two small children are chasing that dog.’ (Warlpiri: Austin & Bresnan 1996)

In (49) two noun phrases ‘two small children’ and ‘that dog’ are not contiguous. Some other elements are interleaved between the head nouns ‘children’ and ‘dog’ and their modifiers ‘small’ and ‘that’. Other permutations of words in the sentence (provided the auxiliary *ka-pala* is always in the second position) can also be uttered spontaneously by a native speaker and the truth-conditional meaning of the sentence does

not change. Austin & Bresnan (1996) have worked out a detailed analysis of non-configurationality in Australian aboriginal languages in LFG framework.[4] Discontinuous noun phrases have also been investigated in detail for many other languages like German (Fanselow & Ćavar, 2002; Kuhn, 1998; Müller, 2004; Roehrs, 2006; van Riemsdijk, 1989), Russian (Gouskova, 2001; Kazenin, 2005; Sekerina, 1997), Greek (Agbayani & Golston, 2005; Nthelitoes, 2004), Dutch (van Hoof, 1997), Ukrainian (Féry *et al.*, 2007), Serbo-Croatian-Bosnian (Boškovič, 2005; Ćavar, 1999), etc.

Different phenomena of discontinuous noun phrases have been discussed in the literature (see for example (Fanselow & Féry, 2006)) and are briefly described in the following subsections. Many examples quoted in these subsections are taken from Fanselow & Féry (2006).

5.4.1.1 Simple and inverted discontinuous noun phrases

If the order of elements in the discontinuous noun phrase is the same as the canonical order in the corresponding continuous noun phrase, then the discontinuous noun phrase is called a simple discontinuous noun phrase, else it is called an inverted discontinuous noun phrase.

Considering the prosodic properties, discontinuous noun phrases are divided into cohesive and non-cohesive discontinuous noun phrases. When the whole discontinuous noun phrase is integrated into a single intonational phrase, then it is a cohesive discontinuous noun phrase and if its two parts are separated into two intonational phrases then it is a non-cohesive discontinuous noun phrase. Simple discontinuous noun phrases are usually cohesive and inverted discontinuous noun phrases are usually non-cohesive. In Ukrainian (Féry *et al.*, 2007) both types of discontinuous noun phrases have been reported.

(50) a. Marija maje **bahato krisel**.
Mary has.got many chairs.Gen.Pl
'Mary has got many chairs.'

b. **bahato** maje Marija **krisel.**

c. **krisel** Marija maje **bahato. (Ukrainian)**

The example sentence (50) from Ukrainian depicts the canonical order of a continuous noun phrase in (50a) and a simple discontinuous noun phrase in (50b) and an inverted discontinuous noun phrase in (50c).

[4] Legate (2002), however, has argued for a configurational analysis of Warlpiri.

5.4.1.2 Extraction from DP

Extraction from DP involves the dislocation of an argument or adjunct of the head noun to the left in the DP. For example in (51), *über Logik* 'about logic' is thematically dependent on the lexical noun *Bücher* 'books'. In (51b), although the DP is discontinuous, the adjunct of the noun is still adjacent to it. In (51c), however, the adjunct *über Logik* is taken out of the DP to the left and hence is an example of extraction.

(51) a. Er hat **viele Bücher über Logik** gekauft.
He has many books on logic bought
'He has bought many books about logic.'

b. **Bücher über Logik** hat er **viele** gekauft.

c. **Über Logik** hat er **viele Bücher** gekauft. (German)

This distinction between extraction and other discontinuous phrases was made by generative syntacticians (Haider, 1985). Extraction as in (51c) is generally explained by movement. Only a maximal projection is posited to move to a pre-auxiliary position. As *über Logik* is the maximal projection of a preposition, it can be moved to the pre-auxiliary position. *Bücher über Logik*, however, is considered a submaximal projection of a noun. The maximal projection of a noun is assumed to have the specifier position filled by the determiner. So (51b) and (51c) are explained by different mechanisms and a distinction is made between (51b) and (51c). Müller (2004) has described various possible analyses for (dis)continuous constituents in German in HPSG with different assumptions and explanations.

5.4.1.3 Quantifier Float

Quantifier Float involves dislocation of the quantified expression away from the noun. This phenomenon has also been explained in terms of movement dependencies in that the DP can move to Spec,TP and the quantifier could be left *in situ* (Déprez, 2003). A further analysis for *all* is made in terms of adverbial quantification as it shares the distributional properties of adverbs like *ever*.

(52) a. **They all** have bought a car.

b. **They** have **all** bought a car.

In (52a) the base generated 'They all' has moved to Spec,TP and in (52b) the quantified part 'all' has been left *in situ* and only *They* has moved to Spec,TP.

McCloskey (2000) has observed another type of quantifier float in the context of *wh*-movement. The quantifier in (53b) is not bound with the subject, rather it is construed with the *wh*- question word referring to the object.

(53) a. **What all** did you get *t* for Christmas?

b. **What** did you get **all** for Christmas? (Irish English)

5.4.1.4 Free Topic structure

In a Free Topic structure, two semantically related elements forming a unique theme become discontinuous in the clause. Usually one element that is more abstract is made the topic and the other more specific element is placed in the canonical position.

(54) **Say-nun** ku-ka **nightingale-man** a-n-ta.
bird-Top he-Nom nightingale-only know-Pres-Dec
'As for birds, he only knows nightingales.' (Korean)

In (54) *bird* and *nightingale* are both semantically related and form a unified theme in the clause, although they are separate from each other in the clause.

5.4.1.5 Extraposition

Extraposition is a phenomenon in which the dependent element of a noun is moved to the right in contrast with extraction where the dependent element is moved to the left.

(55) a. **A man** came in **who had a beard**.

b. **A book** came out **about logic**.

The relative clause in (55a) that describes the noun in the main clause is postposed and in (55b) the PP adjunct of the noun is postposed.

Although extraction, quantifier float, free topic and extraposition are all in some sense discontinuous noun phrases, in the generative framework these are generally distinguished from the absolute discontinuous noun phrases that only involve the separation of the head noun from its determiner, article or an adjective modifying it. Theoretically the phenomenon of a discontinuous noun phrase is licensed only if at least one of the heads involved appears in an A-bar position (Fanselow & Féry, 2006).

5.4.1.6 Clause-level discontinuity in Urdu

Almost all types of discontinuous NPs at clause level mentioned above have also been observed in Urdu. Although clause-level discontinuous NPs are not the main topic here, however, for contrast, evidence for such phrases is provided by the following examples.

(56) a. ندا نے منطق پر ایک کتاب خریدی ہے۔
nɪda=ne **mɑntɪq=pɑr** **ek kɪtab** xɑrid-i hɛ.
Nida=Erg logic=Loc.on one book.F.3Sg buy-Perf be.Pres
'Nida has purchased a book on logic.'

b. منطق پر ندا نے ایک کتاب خریدی ہے۔
mɑntɪq=pɑr nɪda=ne **ek kɪtab** xɑrid-i hɛ.
logic=Loc.on Nida=Erg one book.F.3Sg buy-Perf be.Pres
(Extraction from DP)

(57) a. علی نے بہت آم کھائے۔
ɑli=ne **bɑhʊt am** kʰa-e
Ali.M.3Sg=Erg many mango.M eat-Perf.M.3Pl
'Ali ate many mangoes.'

b. آم علی نے بہت کھائے۔
am ɑli=ne **bɑhʊt** kʰa-e
mango.M Ali.M.3Sg=Erg many eat-Perf.M.3Pl
(Quantifier Float)

(58) a. علی کو آم پسند ہے۔
ɑli=ko **am** pɑsɑ̃d hɛ
Ali.M.3Sg=Dat mango.M liked be.Pres.3Sg
'Ali likes mango.'

b. پھل، علی کو آم پسند ہے.
pʰɑl ɑli=ko **am** pɑsɑ̃d hɛ
fruit.3Sg Ali.M.3Sg=Dat mango.M liked be.Pres.3Sg
'With respect to fruits, Ali likes mangoes' (Free Topic)

(59) a. ایک شغص، جس کی داڑھی تھی، آیا.
ek ʃɑxs **jɪs=ki** **daɽʰi** **tʰi** a-ya
one person.M.3Sg Rel.3Sg=Gen.F beard.F.3Sg be.Past.F.3Sg come-Perf.M.3Sg
'One person, who had a beard, came.'

b. ایک شغص آیا جس کی داڑھی تھی.
ek ʃɑxs a-ya **jɪs=ki** **daɽʰi** **tʰi**
one person.M.3Sg come-Perf.M.3Sg Rel.3Sg=Gen.F beard.F.3Sg be.Past.F.3Sg
'One person came who had a beard.' (Extraposition)

Extraposition with relative clauses and correlatives has been discussed in detail by Dayal (1994) and a relevant discussion can also be found in Dwivedi (1994). Note that the sentence in (57b) is an instance of an inverted discontinuous noun phrase in Urdu. Before moving to the within-constituent discontinuity in Urdu, argument taking adjectives and argument taking nouns are briefly described to provide the necessary background.

5.4.2 NP-internal discontinuity

In NPs, both nouns and their arguments/modifiers can have their own arguments. The discontinuous constituents in NPs occur in Urdu when some argument-taking adjectives modify some argument taking noun or if the specifier of the head noun licenses its own argument in the noun phrase.

Argument-taking adjectives are placed further away from the head noun in comparison with argument-less adjectives. Both the argument of the head noun and the argument of its arguement/modifier can co-occur at the beginning of noun phrases, thus giving rise to discontinuous constituents within noun phrase. Although Urdu

noun phrases have been described in grammar books (see Platts 1967; Schmidt 1999, etc.), the phenomenon of discontinuity within the bounds of noun phrases has not been noticed and discussed before. Consider first rather simple examples of noun phrases in (60)–(61).

(60) a. مقدّمات سے استثنی

mʊqaddɑmat=se ɪstɪsna
court-case.M.3Pl=Abl immunity.M.3Sg
'Immunity from court-cases'

b. سلامتی پر بریفنگ

sɑlamti=pɑr bɑrifiŋg
security.F.3Sg=Loc.on briefing.F.3Sg
'Briefing on security'

c. آرمی چیف سے مطالبہ

armi-cif=se mʊtalbɑh
army-chief.M.3Sg=Abl demand.M.3Sg
'Demand to the army-chief'

(61) a. مقدّمات سے آئینی استثنی

mʊqaddɑmat=se aini **ɪstɪsna**
court-case.M.3Pl=Abl constitutional immunity.M.3Sg
'Constitutional immunity from court-cases'

b. سلامتی پر تفصیلی بریفنگ

sɑlamti=pɑr tɑfsili **bɑrifiŋg**
securit.F.3Sgy=Loc detailed briefing.F.3Sg
'Detailed briefing on security'

c. آرمی چیف سے قانونی مطالبہ

armi-cif=se qanuni **mʊtalbɑh**
army-chief.M.3Sg=Abl legal demand.M.3Sg
'Legal demand to the army-chief'

Example (60) contains just head nouns with a single argument. In (61), however, the head nouns are modified by argument-less adjectives. We see that the argument of the head noun in Urdu is separated from the noun when an adjective modifies the head noun. In English, on the other hand, the adjective modifying the noun is placed prenominally and the complement of the noun comes postnominally and so both remain contiguous to the head noun. A complex example of noun phrases in Urdu with different order of elements is given in (62).

(62) a. صدر کو حاصل مقدّمات سے آئینی استثنی

sɑdɑr=ko$_1$	hasıl$_1$	mʊqaddɑmat=se$_2$	aini	ıstısna$_2$
president=Dat	possessed	court-cases=Abl	constitutional	immunity

‘Constitutional immunity from court-cases possessed by the president’

b. مقدّمات سے صدر کو حاصل آئینی استثنی

mʊqaddɑmat=se$_2$ sɑdɑr=ko$_1$ hasıl$_1$ aini ıstısna$_2$

c. صدر کو مقدّمات سے حاصل آئینی استثنی

sɑdɑr=ko$_1$ mʊqaddɑmat=se$_2$ hasıl$_1$ aini ıstısna$_2$

d. حاصل مقدّمات سے صدر کو آئینی استثنی

* hasıl$_1$ mʊqaddɑmat=se$_2$ sɑdɑr=ko$_1$ aini ıstısna$_2$

e. حاصل صدر کو مقدّمات سے آئینی استثنی

* hasıl$_1$ sɑdɑr=ko$_1$ mʊqaddɑmat=se$_2$ aini ıstısna$_2$

The subscripted numbers in (62) show which arguments belong to which heads. The order of elements in (62a) seems to be canonical where arguments of both noun and adjective are close to their heads. The bracketed structure for (62a) is given as:

[NP[AP[KP sɑdɑr=ko$_1$] hasıl$_1$][KP mʊqaddɑmat=se$_2$] aini ıstısna$_2$]

The bracketed NP shows the logical structure and association of arguments in this noun phrase. The elements marked for case are called case phrases (Butt & King,

2005), hence the case marked arguments of nouns or adjectives are labelled as KP in the bracketing structure. All of the examples in (62a–c) are valid noun phrases of Urdu and are equivalent in meaning. The plausible/canonical order of elements in (62a), although acceptable to some of native speakers of Urdu, is rarely found in newspaper corpus. Instead the orders in (62b–c) are generally found in news corpora, with the latter one being the most common. In (62b) the argument-less adjective is just adjacent to the head noun, then comes the argument-taking adjective with its argument to its left and the argument of the noun is at the left-most edge of the NP. The constituent AP in (62b) is contiguous; however, it becomes non-contiguous in (62c) where arguments of the adjective and the noun are in order on the left; then all the heads follow on the right. The orders in (62d–e) are ungrammatical due to a violation of the head-final constraint in Urdu NPs. More examples of NP-internal discontinuity are shown in (63)–(64), where the head noun has a genitive marked argument and another ablative marked argument.

(63) a. سلامتی پر بریفنگ کا آرمی چیف سے مطالبہ

sɑlamti=pɑr$_1$	bɑrifĩg$_1$=ka	armi-cif=se$_2$	mʊtalbɑh$_2$
security=Loc.on	briefing=Gen	army-chief=Abl	demand

‘The demand to the army chief for a briefing on security’

b. آرمی چیف سے سلامتی پر بریفنگ کا مطالبہ

armi-cif=se$_2$ sɑlamti=pɑr$_1$ bɑrifĩg$_1$=ka mʊtalbɑh$_2$

c. سلامتی پر آرمی چیف سے بریفنگ کا مطالبہ

sɑlamti=pɑr$_1$ armi-cif=se$_2$ bɑrifĩg$_1$=ka mʊtalbɑh$_2$

(64) ملکی سلامتی پر آرمی چیف سے تفصیلی بریفنگ کا قانونی مطالبہ

mʊlki	sɑlamti=pɑr	armi-cif=se	tɑfsili
of-country	security.F.3Sg=Loc.on	army.chief.M.3Sg=Abl	detailed

bɑrifĩg=ka	qanuni	mʊtalbɑh
briefing.F.3Sg=Gen.M.3Sg	legal	demand.M.3Sg

‘The legal demand to the army chief for a detailed briefing on the security of the country’

The argument *barifĩg* 'briefing' of the head noun *mʊtalbah* 'demand' in (63) takes its own case marked complement *salamti=par* 'on security'. All orders in (63a–c) are acceptable to native speakers and the last one is the most common in news corpora. Here, again we see that the argument of the genitive marked argument and the other argument of the head noun stack together on the left edge, and the genitive marked argument and the head noun are lumped together at the right edge. The noun phrase in (64) shows that the argument-less adjectives always have their positions adjacent to the head nouns.

5.4.3 Syntactic explanation

In English, adjectives cannot take complements in their prenominal position. Consider the examples in (65).

(65) a. a proud mother

b. a mother [proud of her daughter]

c. * a [proud of her daughter] mother

d. * a [yellow with age] manuscript (Maling 1983:284)

Emonds (1976) proposed the 'Surface Recursion Restriction' according to which the modifiers themselves cannot be modified in prenominal position. Williams (1982) has formulated the 'Head-Final Constraint' in English: prenominal phrasal modifiers of nouns must be head-final. Due to this constraint no prepositional phrase (PP) can occur before nouns, as shown in (65c–d). The PP *of her daughter* is the complement of the adjective *proud* and *with age* is an adjunct of the adjective *yellow*. Consider some more examples from English in (66).

(66) a. The girl [suitable for the job]

b. * The [suitable for the job] girl

c. * The [for the job suitable] girl

d. The suitable girl for the job

The PP *for the job* is considered as the complement of the adjective *suitable* in (66a) and an adjunct of the head noun in (66d). The construction in (66b) is invalid

because the prenominal adjective is not head-final.[5] Likewise the construction in (66c) is invalid because the complement of an adjective cannot appear before the adjective in English. William's 'Head Final Constraint' on prenominal modifiers also accounts for the German data given in (67).

(67) a. Das Mädchen geeignet [für den Job]
the.Neut girl.Neut suitable for the.Dat job
'The girl suitable for the job'

b. * Das geeignete [für den Job] Mädchen
the.Neut suitable for the.Dat job girl.Neut
'The girl suitable for the job'

c. Das [für den Job] geeignete Mädchen
the.Neut for the.Dat job suitable girl.Neut
'The girl suitable for the job'

d. * Das geeignete Mädchen [für den Job]
the.Neut suitable girl.Neut for the.Dat job
'The girl suitable for the job' (German)

The complement f*ür den Job* 'for the job' of the adjective *geeignet* 'suitable' appears on the right-side of the adjective in (67a–b). The first one is grammatical and the second one is ungrammatical. The same complement appears on the left-side of the adjective in (67c) to meet the Head Final Constraint for prenominal modifiers. The same data can also be accounted for by the Consistency Principle of Giorgi & Longobardi (1991) about (non)-recursion of prenominal modifiers in Romance languages. The principle is stated in (68).

(68) Consistency Principle

An XP immediately expanding a lexical category on the non-recursive side is directionally consistent in every projection.

(Giorgi & Longobardi 1991:98)

The recursive side of a head in this principle is meant by that side where complements appear. Thus if a phrase (XP) occurs on the opposite side of a complement (XP N Complement) then this category would only be expanded to that direction. For German, the recursive side is on the right and if some adjective appears on the left of the head-noun then it will be expanded to the left as is shown in (67b). There are still

[5] Sometimes lexicalized modifiers with prepoitional phrases can appear before nouns in English as in *made-for-tv movies*.

other proposals for prenominal adjecitves where they are treated as heads and adjoin to other heads rather projecting their own structure. Among them are (Bessler, 1992; Sadler & Arnold, 1994; Travis, 1988) for English prenominal adjecitves and (Bouchard, 1997; Lamarche, 1991) are for French prenominal adjectives.

As we have seen in examples of Urdu noun phrases, adjectives can take complements in prenominal position, Emond's Surface Recursion Restriction, therefore, does not apply to Urdu prenominal modifiers. Williams' Head Final Constraint works in Urdu due to default head-last constraint for APs and NPs but does not account for different order of complements in noun phrases. The Consistency Principle of Giorgi and Longobardi cannot be applied to Urdu NPs either, as argument taking adjectives do not appear on the non-recursive side.

In generative syntactic theory, Larson (1997, 1998) has proposed that predicating and adverbial adjectives acquire their surface (prenominal) position in the DP through movement. That proposal has further been extended by Androutsopoulou (2000) to account for pleonastic determiner and prenominal argument taking adjectives in Greek. According to the latter theory the surface position of prenominal adjectives is the result of at least two movements. If the final structure violates Koopman's Generalized Doubly Filled Comp Fileter (Koopman, 1993), it is considered an illegitimate structure. According to this Filter, both the head and the specifier of a projection cannot contain overt material at the end of the derivation.

First movement is the same in all languages in which the AP or extended projection of the AP moves to the specifier of XP1 dominating the NP. The languages only differ in the kind of the second movement. In English, in the second move, the adjectival head undergoes a short head movement from the specifier of XP1 to the head of XP2 immediately dominating XP1. The head noun also moves first to the specifier of XP0, another projection, which is the complement of XP1 and then in a second movement to the head of XP1. The resulting structure for (65c) is shown in Figure 5.6 below. As both the head and the specifier of XP1 are overtly filled in this structure, this structure is ungrammatical.

In languages like Greek, the AP or extended projection thereof moves to the specifier of XP2 and the configuration for a valid phrase of Greek in (69) is generated in Figure 5.7.

(69) i perifani ja tin kori tis mitera
the proud for the daughter hers(CL) mother
'lit. the proud of her daughter mother' (Greek)

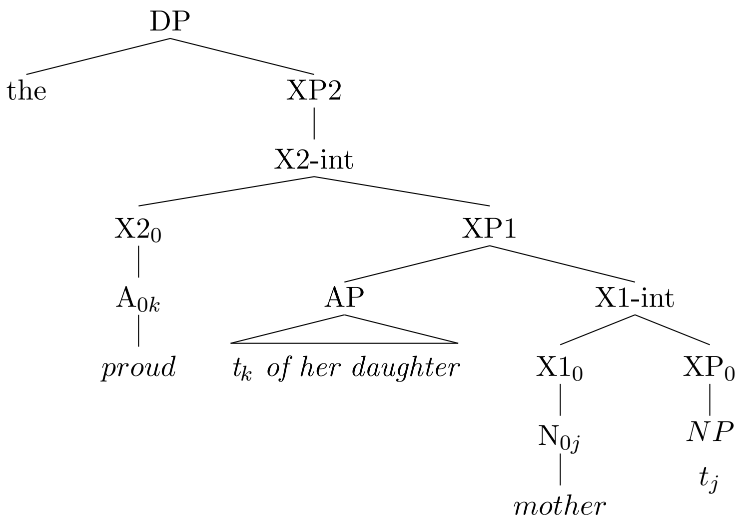


Figure 5.6: English prenominal adjectives

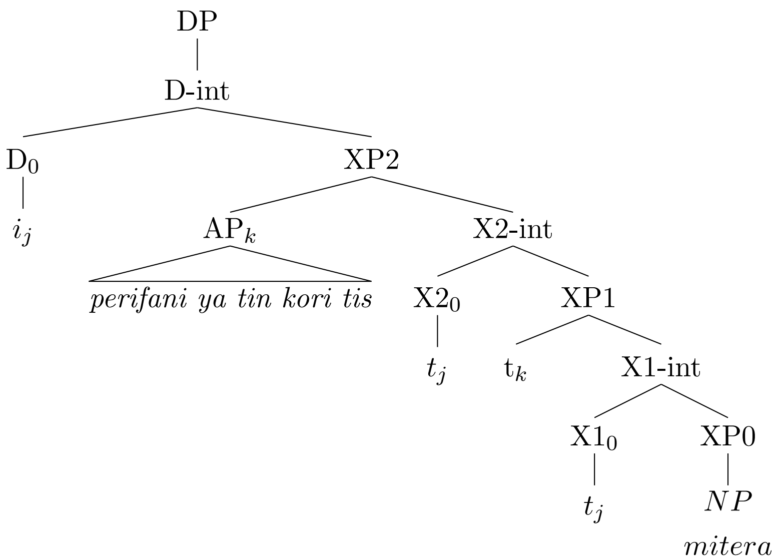


Figure 5.7: Greek prenominal adjectives

The definite determiner in Greek is base generated in the head of XP1 and in English it is generated in the head of the top most projection. Now consider again the Urdu NPs in (70).

(70) a. ? صدر کو حاصل مقدّمات سے استثنی

? sɑdɑr=ko$_1$ hasıl$_1$ mʊqaddɑmat=se$_2$ ıstısna$_2$
president=Dat possessed court-cases=Abl immunity
'Constitutional Immunity from court-cases possessed by the president'

b. مقدّمات سے صدر کو حاصل استثنی

mʊqaddɑmat=se$_2$ sɑdɑr=ko$_1$ hasıl$_1$ ıstısna$_2$

c. صدر کو مقدّمات سے حاصل استثنی

sɑdɑr=ko$_1$ mʊqaddɑmat=se$_2$ hasıl$_1$ ıstısna$_2$

The valid order of different elements in Urdu NPs can also be explained in terms of multiple movements. The first movement as above is of AP to the specifier of XP1. For further steps of movements the complement of head noun moves to successive heads and the complement of adjective moves to the successive specifiers of higher projections both simultaneously or individually. Different movements are illustrated in Figures 5.8–5.10.

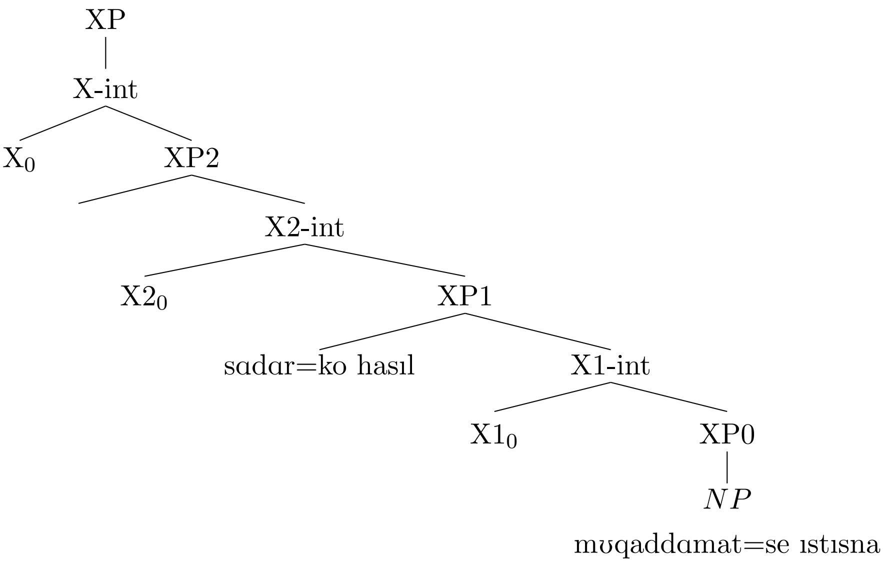


Figure 5.8: Urdu prenominal adjectives 1st movement

As there are no determiners in Urdu, no care is taken to leave heads of different projections to be filled by determiners. After the second movement, Koopman's filter rules out the structure in Figure 5.9. The most frequent structure is shown in Figure 5.10, which is generated after a 3rd movement. If we hold the complement of the adjective in its place and move the complement of a noun in two hops then we can get the second most frequent structure of the noun phrase.

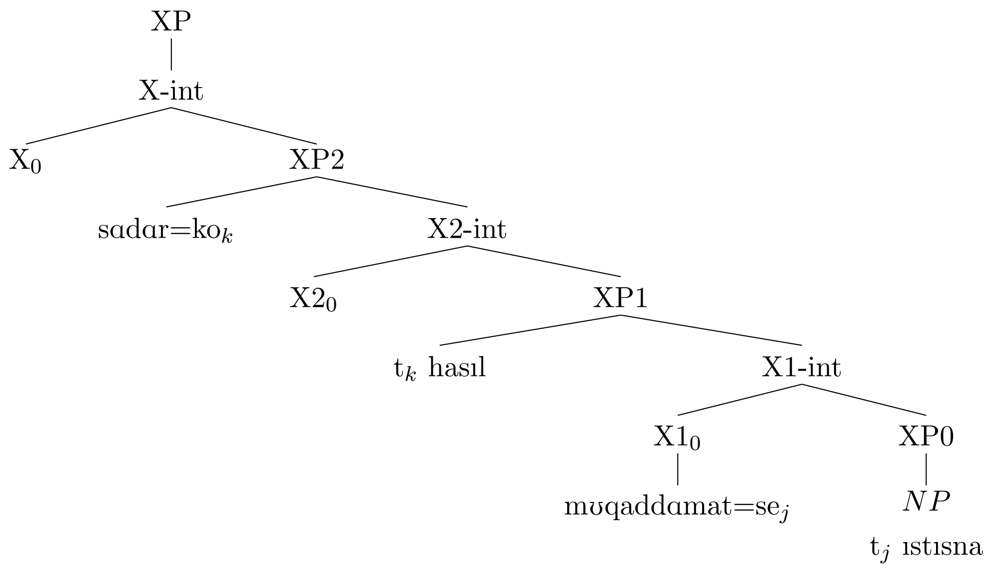


Figure 5.9: Urdu prenominal adjectives 2nd movement

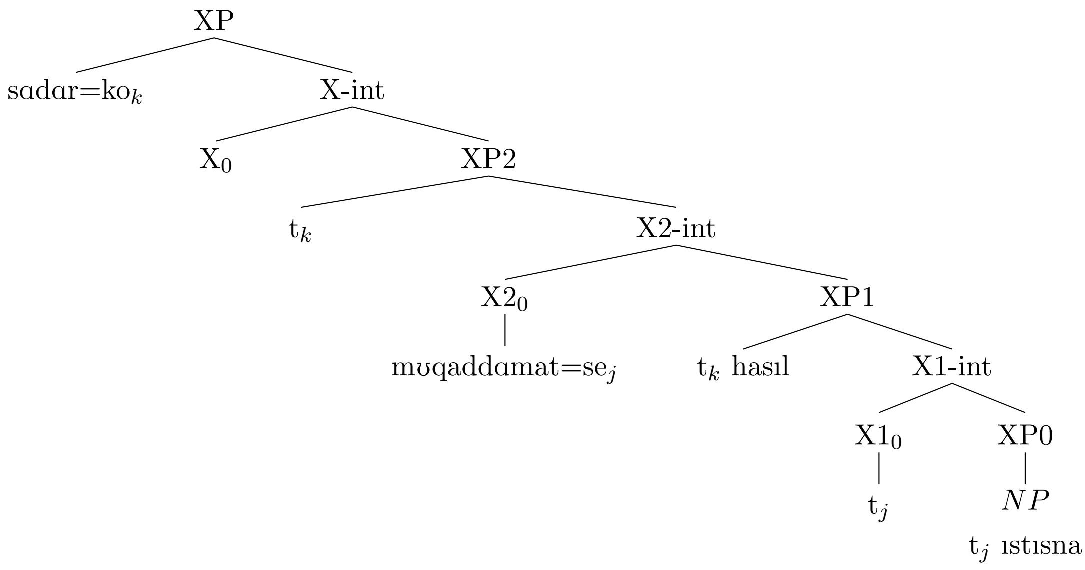
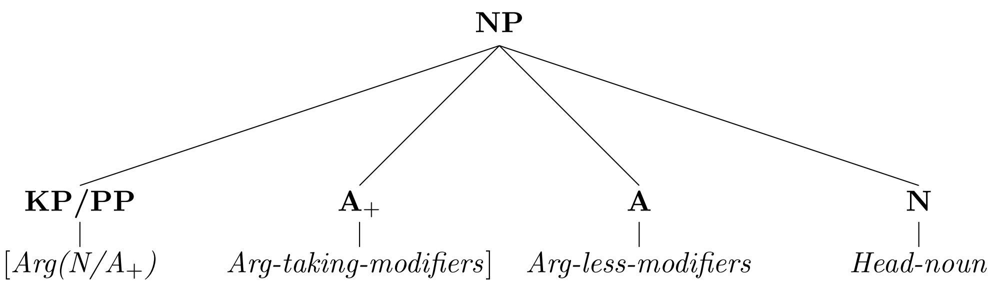


Figure 5.10: Urdu prenominal adjectives 3rd movement

So, the valid orders of different elements in Urdu NPs can be explained by movement across different projections. Koopman's filter rules out invalid orders. In LFG, however, movements cannot be encoded in c-structure, another proposal is provided for c-structure of Urdu noun phrases in general in the next section and it is explained how a correct f-structure of such noun phrases can be generated.

5.4.4 LFG implementation

As discussed above, examples of noun phrases from Urdu news corpora show that the arguments of the head noun and its modifiers/arguments can scramble inside the noun phrase, but that the heads must systematically follow their arguments. Non-continuous APs can appear inside a noun phrase. This evidence of discontinuous constituents within NPs implies a non-hierarchical structure of Urdu NPs. So a flat structure is proposed for Urdu noun phrases (Raza & Ahmed, 2011a). The pattern of NPs in Urdu with the order of different elements is depicted in Figure 5.11. An excerpt from the grammar rules for the implementation of NPs in LFG is given in Figure 5.12.



[Elements in brackets can scramble among each other but the head-final filter is in effect]

Figure 5.11: Word order in Urdu NPs

There are different constraints to be accounted for in generating c-structure and f-structure. For one the head must be final. Secondly the argument-less adjectives must always be contiguous to the head noun they modify. Thirdly allowing different orders of elements in noun phrases.

To model discontinuous XPs at the constituent level within the LFG framework, use of several operators is made. The disjunction notation (|) has been used to assign various functional labels to the KP. The shuffle (,) operator establishes different word orders of the arguments in noun phrase. The $\in$ sign has been used for two different purposes. It is used to add some element to an adjunct set. This is its general use. However, it is also used to assign something nondeterministically to some feature of any member of the adjunct set. Both of its uses appear in the grammar rules for the NP. In the first two lines of the grammar rules in Figure 5.12, $\in$ has been used to assign KP to the OBL function or OBJ-GO function of a member of the adjunct set. Another operator that has been taken advantage of, is the head precedence operator (>h). This operator is used for f-structure precedence and here it is used to make it

$$
\begin{array}{rl}
\text{NP} \rightarrow \text{KP*:} & \{(\uparrow \text{ADJUNCT \$ OBL})=\downarrow \\
 & |(\uparrow \text{ADJUNCT \$ OBJ-GO})=\downarrow \\
 & |(\uparrow \text{OBL})=\downarrow \\
 & |(\uparrow \text{OBJ-GO})=\downarrow\} \\
 & , \text{`}\textit{shuffle operator}\text{'} \\
\text{A}_{+}\text{*:} & \downarrow \in (\uparrow \text{ADJUNCT}) \\
\text{A*:} & \downarrow \in (\uparrow \text{ADJUNCT}) \\
\text{N:} & \uparrow=\downarrow
\end{array}
$$

$$
\begin{array}{rl}
\text{NP} \rightarrow \text{KP*:} & \{(\uparrow \text{ADJUNCT \$ OBL})=\downarrow \\
 & \;(\uparrow \text{ADJUNCT}) >\text{h} (\uparrow \text{ADJUNCT \$ OBL}) \\
 & |\ \ldots\ldots\ldots\} \\
 & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots
\end{array}
$$

Figure 5.12: Grammar Rules

sure that the head will not precede its arguments in the NP, thus implementing the Head-Final constraint. The new rule with this operator is written in the lower part of Figure 5.12. The possible c-structures for (62a)–(62c) are shown in Figure 5.14. In (62c), the hierarchical structure of the AP inside the NP is not possible. So, a flat structure of Urdu NPs is assumed in general. The f-structure representation for each valid instance of (62) is shown in Figure 5.13. In the f-structure we see that logical grouping of different elements is correctly captured.

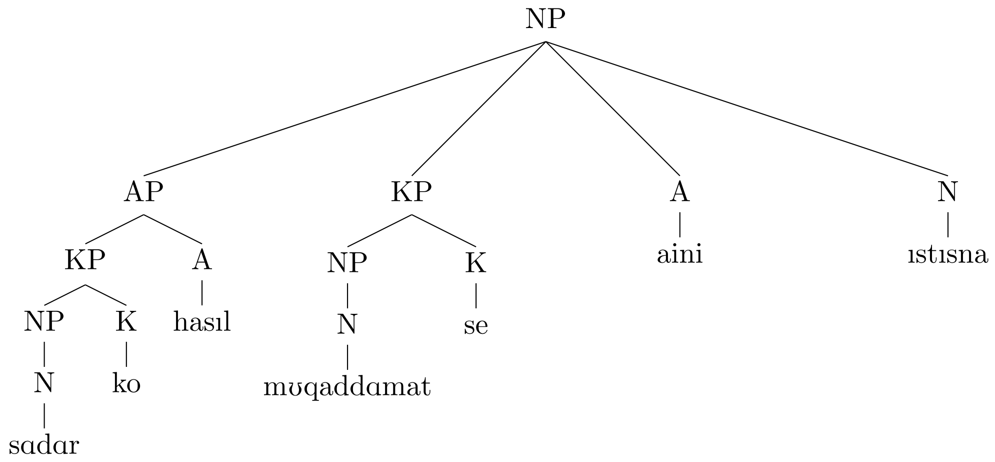
$$
\begin{bmatrix}
\text{PRED} & \text{`ɪstɪsna}\left\langle \text{mʊqaddɑmat} \right\rangle\text{'} \\
\text{OBL} & \begin{bmatrix} \text{PRED} & \text{mʊqaddɑmat} \\ \text{CASE} & \text{abl, GEND masc, NUM pl} \end{bmatrix} \\
\text{ADJUNCT} & \left\{ \begin{array}{l}
\begin{bmatrix} \text{PRED} & \text{`hasɪl}\left\langle \text{sɑdɑr} \right\rangle\text{'} \\ \text{OBL} & \begin{bmatrix} \text{PRED} & \text{sɑdɑr} \\ \text{CASE} & \text{dat, GEND masc, NUM sg, PERS 3} \end{bmatrix} \end{bmatrix} \\
\begin{bmatrix} \text{PRED} & \text{aini} \end{bmatrix}
\end{array} \right\} \\
\text{GEND} & \text{masc, NUM sg}
\end{bmatrix}
$$

Figure 5.13: f-structure representation of (62)

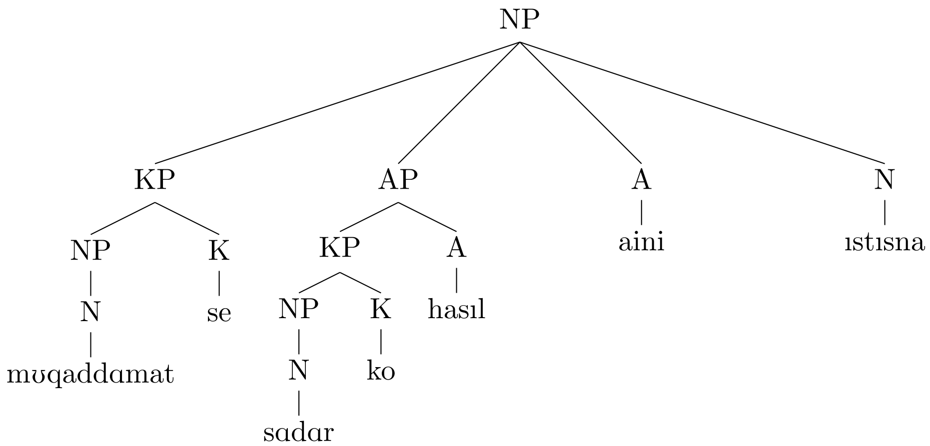
5.5 Conclusion

In this chapter arguments and adjuncts especially the genitive elements in noun phrases were analyzed. The order of different genitival modifiers in Urdu NPs was established. It was shown that the attributive genitives exhibit syntactic distribution similar to adjectives and are placed before adjectives in noun phrases. Some nominlas in Urdu can license exactly two genitive marked arguments and some can take only one genitive marked argument. The discontinuous noun phrases in Urdu were explored in the second part of the chapter.

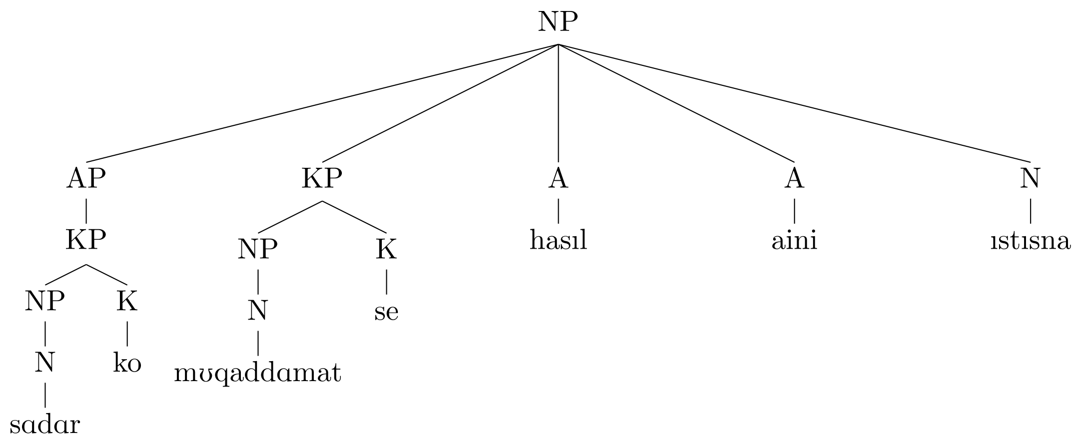
It was shown that discontinuous constituents in Urdu can be found both at the clause level and at the noun phrase level. In NPs, discontinuous constituents arise when an argument-taking noun is modified by an argument-taking modifier or the argument of a noun itself licenses its argument. The argument of the head noun and the arguments of its arguments/modifiers can scramble among each other with the head-final constraint in effect. A syntactic analysis of this phenomenon is provided together with its implementation in the LFG framework.



(a) C-structure for (62a)



(b) C-structure for (62b)



(c) C-structure for (62c)

Figure 5.14: C-structures for the instances in (62)

Chapter 6

Adpositional arguments

Different case marked arguments of verbs were investigated in Chapter 2. Focusing on challenges posed due to the complexity of Urdu case system, an approach to acquire case marked arguments from the raw corpus of Urdu was presented in Chapter 3. The subcategorization information of the verb *ho* and deverbal predicators was investigated in Chapter 4 and Chapter 5, respectively. In this chapter, I analyze different kinds of adpositional participants of the verbs. In addition to the case clitics discussed in previous chapters, there are a variety of adpositions in Urdu. Adpositional phrases, like case phrases, can act either as adverbial modifiers (adjuncts) or as subcategorized-for elements (arguments) of verbs. The internal structure of adpositions is subject to some restrictions. The adposition itself acts as a predicate and requires its complement position to be filled by a case phrase. Not every case phrase is selected by every adposition. In this chapter, the origin of different adpositions in Urdu is described and the structure of adpositional phrases is analyzed. A classification of adpositions is presented based on the type of case phrase complements.

In section 1, adpositions are defined. Section 2 describes the range of spatial adpositons in Urdu and it is argued against the compound adposition analysis of *ke age* 'in front of' and *ke pic*^{h}*e* 'behind' in Butt & King (2005). Section 3 analyzes the structure of spatial adpositions in Urdu in vector space semantics and also in the spirit of the spatial notions worked out by Svenonius (2006, 2007, 2008). A proposal is formulated that encodes Svenonius' notions of spatial expressions in LFG in terms of LexSem features. Nonspatial single word adpositions and compound adpositions are described in section 4. The strucuture of complex adpositions is analyzed in the LFG framework in section 5. Section 6 briefly describes the grammaticalization process in which complex adpositions become simple adpositions. Some more typical adpositions are described in section 7. Classes of adpositions based on alternation of case marking on their complements are given in section 8 and section 9 concludes the chapter.

6.1 What adpositions are

The caseless and tenseless words that express spatial or other relations among entities and events are referred to as adpositions. They differ from verbs as they do not have any aspect or tense and also differ from nouns in that they are not usually pluralized when used as adpositions.

Based on crosslinguistic empirical observations, Svenonius (2007) lists the following typical characteristics of adpositions. Adpositions:

a. Express binary relations between entities (including events)
b. Form a syntactic constituent with a DP complement
c. C-select properties of the complement
d. S-select properties of the complement
e. Project XPs which function as predicates or sentential adjuncts
f. Do not combine with tense or aspect morphology

All of these properties hold for Urdu adpositions. The first property will become clear with the examples in the course of this chapter. Urdu is a free word order language in which major constituents scramble freely in a clause. However, an adposition and its complement cannot be separated in any valid order of the sentence. This phenomenon attests the second property of adpositions in Urdu.

C-selection of a head is the determination of syntactic conditions on a dependent. Just as a verb requires its arguments to appear with a certain case or adposition, so does an adposition require case on its complements. For example, the adposition *pahle* 'before' only takes an ablative marked noun argument. Adpositions can also determine the category of a complement, for example *sat*ʰ 'with' takes only a genitive marked noun, whereas *bayɛr* 'with-out' can also take a perfective participle.

Adpositions may furthermore place semantic restrictions (S-selection) on their complements. For example, *andar* 'inside' only takes a complement which has three dimensional boundaries and some space in between. Adpositions crosslinguistically fill adjunct or complement positions in the projections of verbs and nominals. At least in Urdu, adpositions do not combine with tense or aspect morphology.

6.2 Spatial adpositions

Before progressing to a detailed discussion on true spatial adpositions in Urdu, a brief description of the spatial case markers in Urdu is provided below.

6.2.1 Spatial case markers

Basic prepositions *in, on, to, from* are used in English to encode location and path. To encode such notions in Urdu there are different case clitics. A general description of case clitics was given in Chapter 2. Those clitics which mark nouns for location and path are listed here again in Table 6.1.

Place/Path	**Case Clitic**
Place	*par* 'on' *mẽ* 'in'
Path	*se* 'from' *ko* 'to' *tak* 'up to'

Table 6.1: Spatial case clitics in Urdu

6.2.1.1 Location marking

Two case clitics are used in Urdu to mark locations. The clitic form for adessive case is *par* 'on/at' and for inessive case is *mẽ* 'in'. Examples of their use are given in (1).

(1) a. کتاب میز پر ہے.

kɪtab **mez=pɑr** hɛ
book.F.3Sg table.F.3Sg=Loc.on be.Pres.Sg
'The book is on the table.'

b. بچے سکول میں ہیں.

bacce **sakul=mẽ** hɛ̃
child.Pl school.M.3Sg=Loc.in be.Pres.Pl
'The children are at school.'

In Urdu, a non-verbal noun cannot be modified directly by a locative phrase. Consider, for example, (2)–(3). The noun phrase in (2a) would be ungrammatical in Urdu without the participial adjective *paṛi hu-i* 'lying'. In the English translation, however,

the noun phrase is grammatical even without the participial ‘lying’. Likewise the verbal adjective *mɔjud* ‘present’ is mandatory in (2b) to make the noun phrase grammatically valid with the locative phrase, *sakul=mẽ* ‘in the school’.

(2) a. میز پر پڑی ہوئی کتاب
mez=pɑr *(pɑṛi hu-i) kɪtab
table.F.3Sg=Loc.in *(lie-Perf.F be-Perf.F) book.F.3Sg
‘The book (lying) on the table’

b. سکول میں موجود بچے
sakul=mẽ *(mɔjud) bacce
school.M.3Sg=Loc.in *(present) child.M.Pl
‘The children (present) in the school’

(3) a. سکول میں بچوں کو کتابیں دی گئیں.
sakul=mẽ baccõ=ko kɪtabẽ d-i gɑ-ĩ
school.M.3Sg=Loc.in child.Pl.Obl=Dat book.F.Pl give-Perf.F give-Perf.F.Pl
‘The children were given books in the school.’/
*‘The children in the school were given books.’

b. سکول میں موجود بچوں کو کتابیں دی گئیں.
sakul=mẽ mɔjud baccõ=ko kɪtabẽ d-i gɑ-ĩ
school.M=Loc.in present child.Pl=Dat book.F.Pl give-Perf.F give-Perf.F.Pl
‘The children in the school were given books.’

c. پارلیمان میں آبادی میں اضافے کا مسئلہ اٹھایا گیا.
parliman=mẽ **abadi=mẽ** ɪzafe=ka mɑs’ɑlɑh
parliament.F=Loc.in population.F=Loc.in increase.M=Gen.M issue.M.3Sg
ʊṭ^{h}a-ya gɑ-ya
raise-Perf.M give-Perf.M
‘The issue of increase in population was raised in the parliament.’

In (3a), the locative phrase *sakul=mẽ* ‘in the school’ cannot be considered to be attached to the subject of the sentence. Rather, it is construed that the event of giving

books took place in the school. In (3b), however, the locative phrase is not linked with the main verb of the sentence *de* 'give', instead it is linked with the deverbal adjective *mɔjud* 'present' modifying the noun *bacce* 'children'. In (3c), the first locative phrase is attached at the clause-level and the second locative phrase is the argument of the deverbal noun *ɪzafah* 'increase'. So it can be concluded that in Urdu the locative phrases are attached only with the (de)verbal adjectives or nouns as their arguments.

Absolute locations in Urdu are always overtly marked by case clitics. However, in some Indo-Aryan languages, locations can be unmarked in copula constructions. Consider a contrastive example from Saraiki in (4).

(4) a. علی گھر پر/میں ہے.

ɑli	g^hɑr=*(pɑr/mẽ)	hɛ
Ali.M.3Sg	home.M.Sg=*(Loc.at/Loc.in)	be.Pres.Sg

'Ali is at home.' (Urdu)

b. علی گھر ہے.

ɑli	g^hɑr	he
Ali.M.3Sg	home.M.Sg	be.Pres.Sg

'Ali is at home.' (Saraiki)

The copula construction in (4a) is ungrammatical in Urdu if the noun *g^hɑr* 'home' is not marked with the locative case marker. However, the equivalent construction in (4b) for Saraiki is perfectly grammatical.

6.2.1.2 Path marking

The basic use of three clitics used for marking path in Urdu given in Table 6.1 is illustrated in (5).

(5) a. علی گھر سے آیا.

ɑli	g^hɑr=**se**	a-ya
Ali.M.3Sg	home.M.3Sg=Abl	come-Perf.M.Sg

'Ali came from home.'

b. علی گھر کو گیا.

ɑli g^{h}ɑr=**ko** gɑ-ya
Ali.M.3Sg home.M.3Sg=Loc go-Perf.M.Sg
'Ali went home.'

c. علی دکان تک گیا.

ɑli dʊkan=**tɑk** gɑ-ya
Ali.M.3Sg shop.F.3Sg=Loc go-Perf.M.Sg
'Ali went up to the shop.'

The clitic *se* in (5a) marks source path, while the clitics *ko* and *tak* mark goal/destination paths in (5b) and (5c), respectively.

(6) a. علی باغ سے گزرا.

ɑli baɤ=**se** gʊzr-a
Ali.M.3Sg garden.M.3Sg=Abl pass-Perf.M.Sg
'Ali passed through the garden.'

b. علی گھر گیا.

ɑli g^{h}ɑr gɑ-ya
Ali.M.3Sg home.M.3Sg go-Perf.M.Sg
'Ali went home.'

The clitic *se* is not only used to mark the source path in Urdu, but also the route path as is shown in (6a). The destination path is sometimes not overtly marked with some verbs like *ja* 'go' and *pahʊ̃c* 'reach', etc. in Urdu as is shown in (6b).

(7) a. علی اور ندا دریا کے پل پر جا رہے ہیں.

ɑli ɔr nɪda dɑrya=ke **pʊl=pɑr**
Ali.M.3Sg and Nida.F.3Sg river.M.3Sg=Gen.Obl bridge.F.3Sg=Loc.on
ja rɑh-e hɛ̃
go Prog-Perf.M.Pl be.Pres.Pl
'Ali and Nida are going over the river's bridge.'
'Ali and Nida are going to the river's bridge.'

b. علی اور ندا صنعتی نمائش میں جا رہے ہیں.

ɑli	ɔr	nɪda	sɑn'ɑti	**nʊmaɪʃ=mẽ**
Ali.M.3Sg	and	Nida.F.3Sg	industrial	exhibition.F.3Sg=Loc.in

ja	rɑh-e	hɛ̃
go	Prog-Perf.M.Pl	be.Pres.Pl

'Ali and Nida are walking in the industrial exhibition.'
'Ali and Nida are going to the industrial exhibition.'

Locative phrases are sometimes ambiguous between location and path readings as in (7). In one reading of (7a) and (7b) the locative marked nouns can be construed as places where where Ali and Nida are seen walking and in the other reading the marked nouns can be considered as destinations where Ali and Nida are heading to.

Gehrke (2007) has argued that in English, Dutch and German, prepositions like *in*, *on*, *behind* and *under* are not ambiguous between a locative and directional reading. They only denote places and the meaning of directionality is licensed by other means like additional directional adpositions, resultative verbs, certain movement operations, case and/or contextual or reference axes. In Urdu too, the locative case phrases are interpreted as referring to paths with motion verbs. We see that in both instances of (7) the motion verb *ja* 'go' is used. The directional reading of the instances is possible due to the directional component of the verb *ja* 'go'. However, not all motion verbs license a path meaning with location marked phrases.

6.2.1.3 Stacked markers

Syntactically path structure embeds locative phrases (see e.g. den Dikken 2003; Helmantel 2002; Huybregts & van Riemsdijk 2002; Koopman 1997; Svenonius 2004; van Riemsdijk 1978, 1990). The same is the case with Urdu. A locative marked noun phrase can still be marked with another clitic for path or direction. Consider the instances in (8)–(9), where locative marked phrases are further marked by path clitics.

Only the path clitic *se* 'from' can mark phrases already marked for location by the clitics *mẽ* 'in' or *par* 'on'. The path marking of *mẽ* 'in' marked phrases is illustrated in (8) and that of *par* marked phrases in (9). With verbs like *gʊzar* 'pass' the *se* marked phrases are interpreted as a route as in (8a) and (9a) and with verbs like *nɪkal* 'draw' (8b) and *ʊṭʰa* 'pick up' (9b), a source interpretation obtains.

(8) a. علی باغ میں سے گزرا.
ɑli bɑɣ=**mẽ**=**se** gʊzr-a
Ali.M.3Sg garden.M.3Sg=Loc.in=Abl pass-Perf.M.Sg
'Ali passed through the garden.'

b. علی نے کنویں میں سے پانی نکالا.
ɑli=ne kõvẽ=**mẽ**=**se** pani nɪkal-a
Ali.M.3Sg=Erg well.M.3Sg=Loc.in=Abl water.M.3Sg draw-Perf.F.3Sg
'Ali drew water from the well.'

(9) a. علی پل پر سے گزرا.
ɑli pʊl=**pɑr**=**se** gʊzr-a
Ali.M.3Sg bridge.F.3Sg=Loc.on=Abl pass-Perf.M.Sg
'Ali passed over the birdge.'

b. علی نے میز پر سے کتاب اٹھائی.
ɑli=ne mez=**pɑr**=**se** kɪtab ʊʈʰa-i
Ali.M.3Sg=Erg table.F.3Sg=Loc.on=Abl book.F.3Sg pick-up-Perf.F.3Sg
'Ali picked the book up off the table.'

In Urdu, two place clitics or two path clitics are never stacked together. Crosslinguistically, if two elements are used to encode path meaning, one element is for location marking. The two prepositions *bɪh* 'to' and *ta* 'till' both are used in Persian for marking path (10a)–(10b). In (10c) both prepositions are conjoined to render a path meaning, but the second preposition, here, in fact encodes a location. Although there are other locational prepositions in Persian, sometimes the preposition *bɪh* 'to' is also used to encode location (10d).

(10) a. علی به خانه رفت.
ɑli bɪh xanɪh rɑft
Ali.M.3Sg to home go-Past.3Sg
'Ali went home.' (Persian)

b. علی تا خانه رفت.

ɑli ta xanıh rɑft
Ali.M.3Sg to home go-Past.3Sg
'Ali went home.' (Persian)

c. علی از مسجد تا به خانه رفت.

ɑli ɑz mɑsjıd ta bıh xanıh rɑft
Ali.M.3Sg from mosque.3Sg to at home go-Past.3Sg
'Ali went home from the mosque.' (Persian)

d. میسّر نه گردد به کس این سعادت
به کعبه ولادت به مسجد شهادت

mʊyassɑr nɑh gɑrdɑd bıh kɑs ĩ sɑ'adɑt
available not become.Past.3Sg to anyone this fortune
bıh kɑ'bıh vıladɑt bıh mɑsjıd ʃɑhadɑt
at kabah birth at mosque martyrdom
'(Except Ali) none met such a fortune; birth at Kabah and martyrdom at a mosque' (Persian)

The clitics *se* 'from' and *tak* 'upto' are used to mark source and destination, respectively, of place (11a) or time (11b). These clitics are also used to mark bounds of some set (11c). The phrases marked with such clitics, therefore, represent some bounded place, time or members of some set and can carry a genitive clitic to act as modifier of a head noun (12).

(11) a. علی حاصلوالا سے ملتان (تک) گیا.

ɑli hasılvala=se mʊltan(=tɑk) gɑ-ya
Ali.M.3Sg Hasilwala=Abl Multan(=to) go-Perf.M.3Sg
'Ali went from Hasilwala to Multan.'

b. علی صبح سے شام تک چلا.

ɑli sʊbʊh=se ʃam=tɑk cɑl-a
Ali.M.3Sg morning.F=Abl evening.F=till walk-Perf.M.3Sg
'Ali walked from morning till evening.'

c. ندا نے قلم سے کمپیوٹر تک ہر چیز خریدی.
nıda=ne qɑlɑm=se kɑmpıyuṭɑr=tɑk hɑr ciz xɑrid-i
Nida.F=Erg pen.M=Abl computer.M=till every thing.F purchase-Perf.F
'Nida purchased everything from pen to computer.'

(12) a. حاصلوالا سے ملتان (تک) کا سفر کٹھن تھا.
[hasılvala=se mʊltan(=tɑk)]=ka sɑfɑr kɑṭʰɑn tʰa
[Hasilwala=Abl Multan(=to)]=Gen.M.3Sg journey.M.Sg difficult be.Past.M.Sg
'The journey from Hasilwala to Multan was difficult.'

b. صبح سے شام تک کی پکنک خوشگوار تھی.
[sʊbʊh=se ʃam=tɑk]=ki pıknık xʊʃgɑvar tʰi
[morning.F=Abl evening.F=till]=Gen.F picnic.F.Sg pleasant be.Past.F.Sg
'The picnic from morning to evening was pleasant.'

c. قلم سے کمپیوٹر تک کی اشیاء خریدی گئیں.
[qɑlɑm=se kɑmpıyuṭɑr=tɑk]=ki ɑʃya xɑrid-i gɑ-ĩ
[pen.M.Sg=Abl computer.M.Sg=till]=Gen.F thing.F.Pl purchase-Perf.F go-Perf.F.Pl
'Things ranging from a pen to a computer were purchased.'

If the source point of some place or time or the initial boundary of some set is unspecified or understood, only the destination or the other boundary is encoded by marked noun with a *tak* clitic and that further can carry the genitive marker to modify some other noun (13).

(13) a. ملتان (تک) کا سفر کٹھن تھا.
mʊltan(=tɑk)=ka sɑfɑr kɑṭʰɑn tʰa
Multan(=to)=Gen.M.3Sg journey.M.3Sg difficult be.Past.M.Sg
'The journey to Multan was difficult.'

b. اب تک کی تحقیقات تسلّی بخش ہیں۔

ab=tɑk=ki tɑhqiqat tɑsɑlli-bɑxʃ hɛ̃
now=till=Gen.F investigation.F.3Pl satisfactory be.Pres.Pl
'Investigations made till now are satisfactory.'

c. سوئی تک (کی چیز) ندا کو جہیز میں دی گئی۔

sui=tɑk(=ki ciz) nɪda=ko jɑhez=mẽ
needle.F.Sg=till(=Gen.F thing.F.Sg) Nida=Dat dowry.M=Loc.in
d-i gɑ-i
give-Perf.F.Sg go-Perf.F.Sg
'Nida was given things down to a needle in her dowry.'

We have thus seen that stacking of clitics is possible in two cases: 1) location marked nouns further marked by the path clitic *se* to encode the source/via path, 2) *tak* marked nouns further marked by genitive clitics to act as modifiers. Otherwise no overtly marked nouns can be marked further by another clitic.

6.2.2 Spatial postpositions

The postpositions which in combination with an NP (object) are used to denote either location or path are called spatial adpositions. A list of spatial postpositions used in Urdu is given in Table 6.2.

In addition to the above spatial postpositions there is one comitative postposition in Urdu *sat*ʰ 'with', which is also sometimes used for spatial vicinity. Other postpositions *baɣal mẽ* 'in armpit', *pahlu mẽ* 'in facet' and *barbar mẽ* 'in equal' are also used to encode the meaning of spatial closeness. The adposition *darmɪyan* is originally a combination of Persian basic preposition *dar* 'in' and other word *mɪyan* 'center' and is lexicalized in Urdu as one word.[1] In addition to *taraf* 'side/direction', two more postpostions *sɪmat* 'side/direction' and *janɪb* 'side/direction' are also used. The use of spatial adpositions is exemplified in (14).

[1]This is also the case with the words *dar-peʃ* 'faced' and *dar-pɛ* 'yielding/after', which are each used as one word in Urdu.

Nr.	Postposition	Meanings
(i)	ɑndɑr	‘inside’
(ii)	bahɑr	‘outside’
(iii)	age	‘in front of’
(iv)	picʰe	‘behind’
(v)	upɑr	‘above/over’
(vi)	nice	‘under’
(vii)	tɑle	‘under’
(viii)	samne	‘against’
(ix)	gɪrd/ɪrd gɪrd	‘around’
(x)	nɑzdik	‘near’
(xi)	qɑrib	‘near’
(xii)	pas	‘near’
(xiii)	bic	‘between’
(xiv)	dɑrmɪyan	‘between’
(xv)	kɪnare	‘side’
(xvi)	tɑrɑf	‘side/direction’
(xvii)	par	‘other side/across’

Table 6.2: Spatial postpositions in Urdu

(14) a. علی دکّان کے آگے رکا.

ɑli dʊkan=ke **age** kʰɑṛa hɛ
Ali.M.3Sg shop=Gen.Obl front standing be.Pres.3Sg
‘Ali is standing in front of the shop.’

b. کرسی کے نیچے بلّی ہے.

kʊrsi=ke **nice** bɪlli hɛ
chair.F.3Sg=Gen.Obl under cat.F.3Sg be.Pres.3Sg
‘There is a cat under the chair.’

The use of genitive marker is discussed in section 5.2.1. In Urdu grammar books (e.g. Schmidt 1999) and the linguistic literature (e.g. Butt & King 2005) *ke age* and *ke nice* are generally considered to be compound postpositions. These spatial postpositions

always appear with the genitive, that is why the compound postposition analysis was proposed. I argue that these in fact are not compound adpositions.

Many morphemes used for spatial postpositions in Urdu originally are nouns and the form of the genitive case used with them is due to the gender of the original noun (Platts, 1967). So, the genitive oblique form *ke* in them should be considered as a case marker rather a part of a compound postposition. Arguments for the analysis of *ke age* as two separate entities, namely, a genitive case clitic and a spatial adposition that was once a noun, come from: the other form of genitive with the first person and second person pronouns, the placement of emphatic clitics, the general separability of the genitive and the adposition, use of another case marker in place of the genitive, omissibility of the genitive marked NP, coordination, postposition and agreement. These data data are presented in the remainder of this section.

(i) With the first person and second person pronouns the genitive appears as part of pronouns (15b), that is the genitive form *ke* is not used.

(15) a. کرسی کے نیچے

kʊrsi= **ke** **nice**
chair.F.3Sg=Gen.Obl under
'under the chair'

b. میرے نیچے

mere **nice**
1Sg.Gen.Obl under
'under me'

(ii) The emphatic marker *hi* can intervene between the case marker *ke* and the adposition (16).

(16) کرسی کے ہی نیچے

kʊrsi= **ke** hi **nice**
chair.F.3Sg=Gen.Obl Emph under
'really under the chair'

(iii) A measure/description term can come between the case marker *ke* and the adposition (17). In addition to measure phrases, spatial adpositions can also be modified by special adverbs like *bahʊt* 'much', *bɪlkʊl* 'exactly', and *ɛn* 'exactly'.

(17) زمین کے ۳ فٹ نیچے
zɑmin=**ke** 3 fʊʈ **nice**
ground.F=Gen.Obl 3 foot under
'three feet under the ground'

(iv) The complement of an adposition can optionally be marked by an alternate case marker (18).

(18) زمین سے ۳ فٹ نیچے
zɑmin=**se** 3 fʊʈ **nice**
ground.F=Abl 3 foot under
'three feet under the ground'

(v) The *ke* marked complement can be postposed (19).

(19) ۳ فٹ نیچے زمین کے
3 fʊʈ **nice** zɑmin=**ke**
3 foot under ground.F=Gen.Obl
'three feet under the ground'

(vi) The clitic form *ke* can be distributed over the coordinated adpositions (20).

(20) کرسی کے اوپر اور نیچے
kʊrsi=**ke** **upɑr** ɔr **nice**
chair.F.3Sg=Gen.Obl over and under
'over and under the chair'

(vii) The genitive marked complement, which semantically is the ground, can sometimes be omitted when understood (21). Naked spatial adpositions in Urdu behave like adverbs. The category P has been realized for both dressed and naked adpositions of

Hungarian by Marácz (1989), however, naked Ps have been termed as adverbs by Kiss (2002). I would follow the first analysis. In case of naked adpositions, the ground is implicitly understood.

(21) علی، (گھر کے) باہر آؤ!
ɑli (g^{h}ɑr=ke) **bahɑr** a-o
Ali.M.Sg home.M.3Sg=Gen.Obl outside come-Imper.2Sg
'Ali, come out (of home)!'

(viii) The feminine form of a genitive marker is used when the adposition originates from a feminine noun (22). The agreement of genitive strengthens the analysis that the genitive form with these adpositions should be considered as a case marker rather than the part of compound adposition.

(22) علی بازار کی طرف چلا.
ɑli bazar=**ki** **tɑrɑf** cɑl-a
Ali.M.Sg market.M.3Sg=Gen.F side head-Perf.M.3Sg
'Ali headed towards the market.'

As already mentioned, many spatial adpositions in Urdu are originally nouns. Many of them are not used as nouns anymore. Some of them do continue to be used as nouns, but their syntactic distribution is different when used as adpositions and when used as nouns. Spatial adpostions give a locational sense usually without being overtly marked by a case clitic (23) in comparison with nouns.

(23) دکّان کے آگے (* پر/میں)
dʊkan=ke **age**(=*pɑr/mẽ)
shop.F.3Sg=Gen.Obl front(=Loc.on/in)
'In front of the shop'

In modern Urdu the spatial adposition morpheme *age* 'front' is not used as a noun (24).

(24) a. علی دکّان کے آگے رکا.

ɑli dʊkan=ke age rʊk-a
Ali shop.F.3Sg=Gen.Obl front stop-Perf.M.3Sg
‘Ali stopped **in front of** the shop.’

b. جہاز کے آگے کو آگ لگ گئی.

* jɑhaz=ke **age**=ko ag lɑg gɑ-i
aeroplane.M.3Sg=Gen.Obl front=Acc fire.3F.Sg hit go-Perf.F.3Sg
‘**The front** of the aeroplane caught fire’

c. جہاز کے آگے میں آگ لگ گئی.

*jɑhaz=ke **age**=mẽ ag lɑg gɑ-i
aeroplane.M.3Sg=Gen.Obl front=Loc fire.F.3Sg hit go-Perf.F.3S
‘**The front** of the aeroplane caught fire’

In contemporary Urdu, for the meaning of noun ‘front’, the NP *agla hɪssah* is used instead (25).

(25) a. جہاز کے اگلے حصّے کو آگ لگ گئی.

jɑhaz=ke **ɑgle hɪsse**=ko ag lɑg gɑ-i
aeroplane.M.3Sg=Gen.Obl front part.M.3Sg=Acc fire.F.3Sg hit go-Perf.F.3Sg
‘**The front** of the aeroplane caught fire.’

b. جہاز کے اگلے حصّے میں آگ لگ گئی.

jɑhaz=ke **ɑgle hɪsse**=mẽ ag lɑg gɑ-i
aeroplane.M.3Sg front part.M.3Sg=Loc.in fire.F.3Sg hit go-Perf.F.3Sg
‘**The front** of the aeroplane caught fire.’

In contrast with nouns, the adjectival modification and pluralization of spatial adpositions is not possible in Urdu except for the postposition *taraf* ‘side’ which can be modified by demonstratives *is* ‘this’, *us* ‘that’ or the adjectives *daĩ* ‘right’ and *baĩ* ‘left’, or by quantifiers like *donõ* ‘both’, *carõ* ‘all four’, etc. Its plural form *atraf* ‘sides’ can also be used due to ideosyncratic semantics of this spatial adposition.

6.2.3 Persian spatial prepositions in Urdu

In Persian, there are two basic prepositions for location: *bar* 'on' and *dar* 'in' and three for path: *az* 'from', *bɪh* 'to' and *ta* 'up to' in Persian. These are rarely lexically contentful in Urdu. Instead, they are used grammatically in some adverbial constructions in Urdu. In contrast to these basic Persian prepositions, some other Persian prepositions are used in a lexical meaning in Urdu. Examples of Persian prepositions used in Urdu are given in Table 6.3. It should be noted that Persian *bɪh* is pronounced as *bah* in Urdu.

Presian preposition	**Examples found in an Urdu corpus**
bar 'on'	*bar vaqt* 'in time', *bar mahal* 'in the nick of time/fitting'
dar 'in'	*dar haqiqat* 'in fact', *dar pardah* 'behind the curtain/hidden'
az 'from'	*az sar-e nɔ* 'from scratch', *az xʊd* 'by itself'
bɪh/bah 'to'	*ba-zahɪr* 'apparently', *ba-ja* 'right/accurate/in place"
ta 'up to'	*ta hal* 'till now', *ta qɪyamat* 'till resurrection'
pas 'behind'	*pas-e gardan* 'behind the neck', *pas-e pardah* 'behind the curtain'
bala 'above'	*bala-e mɪmbar* 'over the pulpit', *bala-e sar* 'over the head' '
zer 'under'	*zer-e zamin* 'under ground', *zer-e ab* 'under water'

Table 6.3: Use of Persian spatial prepositions in Urdu

The preposition *zer* is more frequently used in Urdu as compared to other prepositions listed in the table. Some prepositional phrases of Persian, involving the demonstratives *in/ĩ* 'this' or *an/ã* 'that' in them, are used in Urdu as multiword colloquial expressions, for example, *dar in ɪsna* 'in the mean time', *bɪna bar in* 'based on this', *ɪlavah bar in* 'in addition to this', *qabl az in* 'before this', *peʃ az in* 'before this', *ba'd az in* 'after this' and *ba'd az an* 'after that'.

In this section, a range of spatial adpositions was discussed. There are some clitics which are used to mark locations. Many spatial adpositions used in Urdu are originally nouns. It was argued that the genitive oblique form used with a spatial postposition is not the part of the postposition. Rather it marks the object of the postposition for genitive case. The use of Persian prepositions in Urdu was given for the sake of completeness.

6.3 Analysis of spatial adpositions

6.3.1 Vector Space Semantics

To explain the modification of PPs with measurement terms, PPs have been interpreted as a set of vectors, rather than as a set of points or mereological portions of space (Zwarts, 1995, 2000; Zwarts & Winter, 1997). Special terms and extended expressions are interpreted in the context of three dimensional *vector space*. A vector is a directed line segment between points in space.

Spatial adpositions are seen as vector spaces projected from the Ground. The measure phrases define a subset of these vectors. For example, *mez ke upar* 'above the table' defines a space of vectors emanating from the upper plane of the table (see Figure 6.1). And the measure phrase in *mez ke 2 miṭar upar* '2 meters above the table' selects the subset of that space constituted by vectors of 2 meters length each, as shown on the right hand side of Figure 6.1.

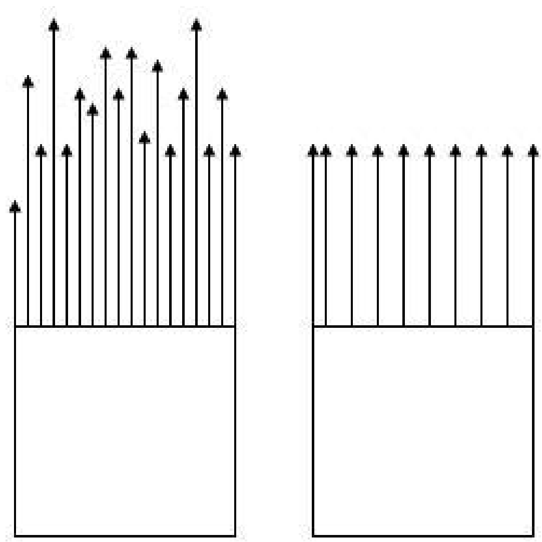


Figure 6.1: Vector Space

Initially, spatial adpositions were analyzed as denoting a general region of space relative to the Ground. However, this analysis failed to properly explain the measure phrases in PPs. On the other hand, if the region is considered to consist of vectors originating from the reference plane, then how the modifiers of measure phrases work can be explicated clearly. The explanation of the PPs in (26) is given in (27).

(26) a. مکان کے آگے

makan=ke **age**
house.M.3Sg=Gen.Obl front
'in front of the house.'

b. مکان کے ۵ میٹر آگے

makan=ke 5 miṭar **age**
house.M.3Sg=Gen.Obl 5 meter front
'five meters in front of the house'

c. مکان کے سیدھا آگے

makan=ke sid^{h}a **age**
house.M.3Sg=Gen.Obl straight front
'straight in front of the house'

d. مکان کے بالکل آگے

makan=ke bılkʊl **age**
house.M.3Sg=Gen.Obl just front
'just in front of the house'

(27) a. [[in front of the house]] = {v | v originates at the front plane of the house and extends beyond it}

b. [[5 meters in front of the house]] = {v | v originates at the front plane of the house and extends beyond it and v has a measure of 5 meters}

c. [[straight in front of the house]] = {v | v originates at the front plane of the house and extends beyond it and v is perpendicular to the front plane}

d. [[straight in front of the house]] = {v | v originates at the front plane of the house and extends beyond it and v has a measure approximately equal to zero}

The vector space system has the following formal components (for details see Winter 2001; Zwarts & Winter 1997):

1. A vector space V over the real numbers **R**.
2. An addition operation + on elements of V.
3. A zero element 0 that satisifes $v + 0 = v$ for all $v \in V$.
4. An opposite element -v for each $v \in V$, such that v + (-v) = 0.
5. A scalar multiplication operator between real numbers in **R** and elements in V such that for all s ∈ **R** and $v \in V$, s $v \in V$.
6. A norm function || from V to non-negative elements of **R**.

Zwarts (2000) defined the place vector *place(v, x)* as the vector v originating at y. The analysis of some locative PPs in terms of place vectors is given in (28).

(28) a. [[inside NP]] = {v | *place*(v, [[NP]]) ∧ |v| ≺ 0}

b. [[outside NP]] = {v | *place*(v, [[NP]]) ∧ |v| ≻ 0}

c. [[in front of NP]] = {v | *place*(v, [[NP]]) ∧ |FRONT(v)| ≻ |⊥ FRONT(v)|}

d. [[behind NP]] = {v | *place*(v, [[NP]]) ∧ |-FRONT(v)| ≻ |⊥ -FRONT(v)|}

e. [[above/over NP]] = {v | *place*(v, [[NP]]) ∧ |VERT(v)| ≻ |⊥ VERT(v)|}

f. [[below/under NP]] = {v | *place*(v, [[NP]]) ∧ |-VERT(v)| ≻ |⊥ -VERT(v)|}

The *axis functions* VERT, FRONT, etc., their inverses (-VERT, etc.) and their orthogonal complements (⊥ VERT, etc.) are added to deal with directional adpositions. Different measure or description phrases modifying the PPs actually represent more specified set of vectors as is shown below:

(29) a. [[5 meters]] = {v | |v| = 5m}

b. [[straight]] = {v | VERT(v)}

c. [[just]] = {v | |v| ≈ 0}

This is how the vector space system (VSS) explains the spatial adpositions and their modification by a measurement or descriptive phrases. Stacked adpositions in Urdu can also be anlyzed in this system. The position of an object can be interpreted to be a vector resulting from the summation of two vectors entailed by two stacked adpositions. Consider the PPs in (30).

(30) a. دروازے کے اوپر دائیں طرف
darvaze=ke upar daĩ taraf
door.M.3Sg=Gen.Obl above right side
'above the door and to the right of it'

b. دروازے کے دائیں طرف اوپر
darvaze=ke daĩ taraf upar
door.M.3Sg=Gen.Obl right side above
'to the right of the door and above it'

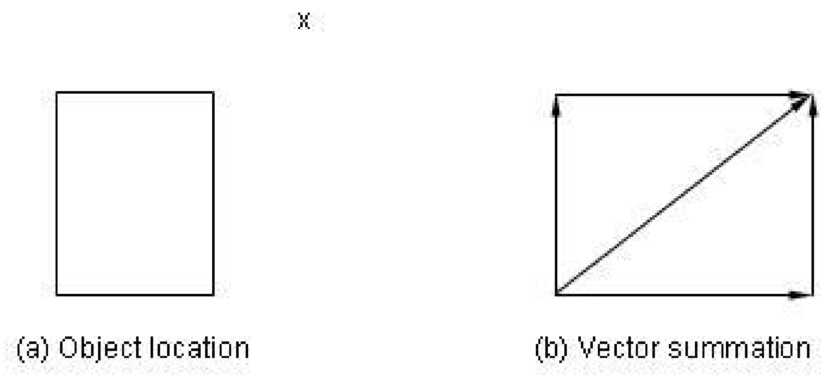


Figure 6.2: Interpretation of spatial adpositions

The two-dimensional object position entailed by the spatial adpositions in (30) is illustrated in Figure 6.2 in the context of VSS. The summation of upward and rightward vectors for two spatial adpostions in (30a) results in a diagonal vector which shows the position of the object with reference to the door. The same resultant vector is produced if the vectors are added in the reverse direction which correspond to the order of spatial adpositions in (30b). The position of the object in (30) is described by stacking two adpositions. To describe the position of an object in two dimensions either an adverbial modification of a single PP is made or two prepostional phrases are coordinated.

6.3.2 Svenonius' hierarchy and notions of space

Svenonius (2007, 2008) has examined a set of crosslinguistic data with respect to PPs in order to arrive at a universal structure of PPs. He proposes a rather fixed hierarchy of different functional heads for spatial PPs:

p - Deg(ree) - Deix(is) - Loc - Ax(ial)Part - K - DP

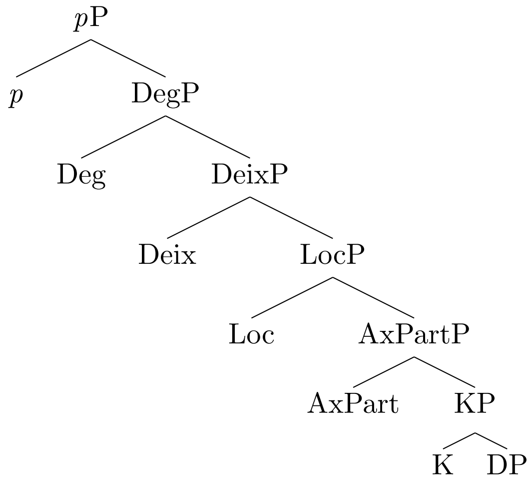


Figure 6.3: Svenonius' hierarchy of spatial adpositions

In this hierarchy, different functional heads are recursively nested. The functional heads do not always have to be realized. They could also be null. In Figure 6.3, *p* is the prepositional head. The Deg head corresponds to the measure phrases in PPs. The AxParts correspond to so-called "Axial Parts" of an object and are described by Jackendoff (1996) as follows:

"The "axial parts" of an object – its *top*, *bottom*, *front*, *back*, *sides*, and *ends* – behave grammatically like parts of the object, but, unlike standard parts such as a *handle* or a *leg*, they have no distinctive shape. Rather, they are regions of the object (or its boundary) determined by their relation to the object's axes. The up-down axis determines top and bottom, the front-back axis determines front and back, and a coplex set of criteria distinguishing horizontal axes determines sides and ends."

(Jackendoff 1996:14)

The head K in the structure is for the case of complements as the complements of PPs are usually marked for case. The Deix head is usually filled by demonstratives which just express different degrees of proximity to a deictic center (Svenonius, 2006). For example, the English PP '5 meters in front of the house' is analysed in an hierarchical structure, as shown in Figure 6.4.

Although almost all notions of spatial expressions worked out by Svenonius are found crosslinguistically, the hierarchy of these notions can vary from language to lan-

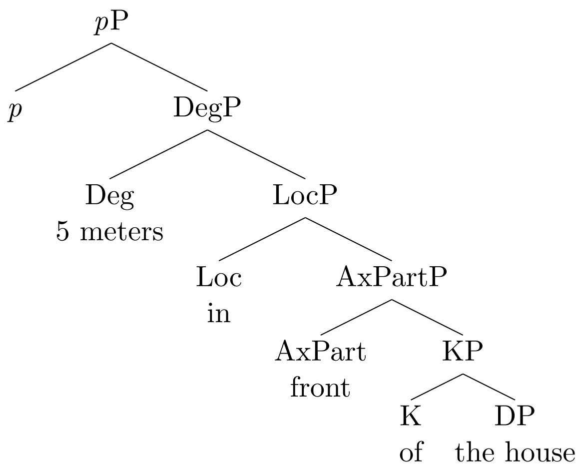


Figure 6.4: English PP structure

guage. For example, in Persian, the Loc head is above Deg and Deix (proximal/distal marker) heads (Pantcheva, 2006, 2008; Svenonius, 2008), as shown in (31).

(31) در ۵ متری این جلوی خانه
dɑr 5 mɪtrɪ-ye in jɪlu-ye xanɪh
at 5 meter-Ez Prox front-Ez house
'here, five meters in front of the house' (Persian)

The list of spatial adpositions in Table 6.2 is in fact a list of AxParts in Urdu. The following hierarchy can be observed for different spatial notions in Urdu spatial postpositons.

Urdu PP: NP - K - Deg(ree) - Deix(is) - Ax(ial)Part - Loc - *p*

The evidence for the observed hierarchy in Urdu can be seen in (32) below.

(32) a. مکان کے ۵ میٹر آگے

mɑkan=ke	5	mɪṭɑr	age
house.M.3Sg=Gen.Obl	5	meter	front

‘five meters in front of the house’

b. مکان کے بیچ (میں)

mɑkan=ke	bic	(mẽ)
house.M.3Sg=Gen.Obl	center	(in)

‘in the middle of the house’

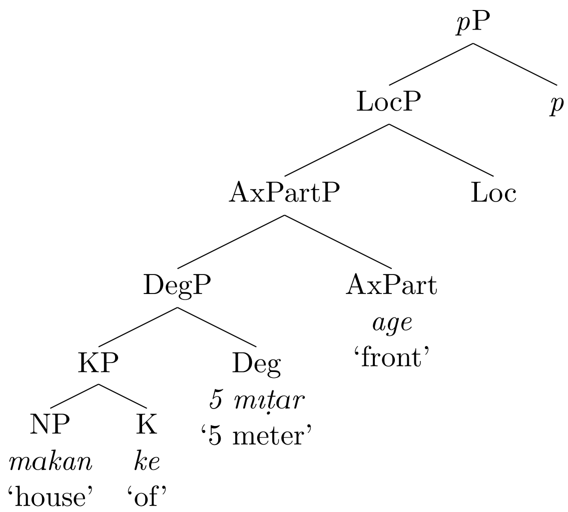


Figure 6.5: Urdu PP structure

The structure for (32a) is displayed in Figure 6.5. If we take the Deg head out of the tree structure (for a moment), then we see the structure of Urdu postpositions just becomes the mirror image of English prepositions. So, ignoring the position of Deg/Deix heads, Svenonius’ hierarchy of spatial notions indeed seems to work crosslinguistically.

For a Deix example in Urdu, consider the PP in (33a) below. The word *par* in Urdu means ‘the other side of some river, stream, etc’. When used as a postposition it means ‘across’ and the demonstrative *ʊs* in (33) is the Deix head which expresses that the region referred to is far across the river rather than just across the river.

(33) a. دریا کے (اس) پار
dɑrya=ke (ʊs) par
river.M.3Sg=Gen.Obl Dist across
'(far) across the river'

b. دریا کے (اس) پار سے
dɑrya=ke (ʊs) par=se
river.M.3Sg=Gen.Obl Dist across=from
'from (far) across the river'

In Svenonius' hierarchy, the path head of spatial adpositions is at the top of all place components. This is the case for Urdu as well, as in (33b).

6.3.3 LFG Model of spatial expressions in terms of Lex-Sem features

As mentioned in the previous section, the same notions of spatial expressions are found crosslingistially, although some elements are placed at different positions in the hierarchy. So, the c-structure of PPs in the LFG paradigm can vary from language to language. However, so far the f-structure for prepositional phrases should look parallel due to the same functional elements in PPs across languages. However, the f-structure of corresponding PPs in the existing ParGram grammars (see section 1.4) of different languages are not parallel. For example, the f-structures of English PP *5 meters behind the house* and its German equivalent *5 Meters hinter dem Haus* are shown in Figures 6.6 and 6.7.

```
[PRED       'meter'
 SPEC       [NMBER [PRED '5'
                    NUMBER-TYPE  card]]
 ADJUNCT    {[PRED   'behind ⟨house⟩'
              PSEM   [{loc}]
              OBJ    [PRED  'house'
                      SPEC  [DET [PRED      'the'
                                  DET-TYPE  def]]
                      PERS  3, NUM sg, CASE obl]
              PTYPE  sem]}
 STMT-TYPE  header, PERS 3, NUM pl]
```

Figure 6.6: f-structure for English PP

$$\begin{bmatrix} \text{PRED} & \text{`hinter}\left\langle\text{Haus}\right\rangle\text{'} \\ \text{ADJUNCT} & \left\{ \begin{bmatrix} \text{PRED} & \text{`Meter'} \\ \text{SPEC} & \begin{bmatrix} \text{NMBER} & \begin{bmatrix}\text{PRED} & \text{`5'} \\ \text{NUMBER-TYPE} & \text{card}\end{bmatrix}\end{bmatrix} \\ \text{NUM} & \text{pl, GEND masc, CASE gen}\end{bmatrix}\right\} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`Haus'} \\ \text{SPEC} & \begin{bmatrix} \text{DET} & \begin{bmatrix}\text{PRED} & \text{`the'} \\ \text{DET-TYPE} & \text{def}\end{bmatrix}\end{bmatrix} \\ \text{PERS} & \text{3, NUM sg, GEND neut, CASE dat}\end{bmatrix} \\ \text{STMT-TYPE} & \text{header, PTYPE sem, PSEM loc}\end{bmatrix}$$

Figure 6.7: f-structure for German PP

In the English ParGram grammar, *meter* is the head and *behind the house* is a prepositional adjunct while in the German ParGram grammar, *hinter* (preposition) is the head, *Haus* is its object and the measure phrase *5 Meters* is the adjunct. The German f-strucutre seems to be more reasonable, because the AxPart is chosen as the head of the prepositional phrase. The AxPart is the key element in such phrases in the sense that some other elements like Deg(ree) and Deix(is) in Svenonius' hierarchy cannot be licensed if there is no AxPart. So, it is proposed that in such phrases the AxPart should in principle be chosen as a head.

Drawing on the functional heads of Svenonius' hierarchy, features are proposed for different place components (Raza & Ahmed, 2011b). The lexical semantics of spatial components can be included in the f-structure in the form of features/attributes under a LEX-SEM label which already exists independently in the Urdu PARGRAM grammar. The proposed features for different components of PLACE/PATH are listed in Table 6.4. A single token of PPs can have one or more features. Distribution of features over tokens of the phrase *5 meters behind the house* would be as follows:

The bare form of AxParts as *fron*t in *in front* would not have a LOC feature as it would be carried by the preposition *in*. In different languages notions of spatial expressions can be encoded either morphologically or syntactically. This is illustrated by the following examples:

- from in front (English)
  Src | Loc | AxPart (3 tokens, 3 features)

PLACE/PATH	Feature	Value
PLACE	AxPart	back, front, etc
	Loc	+
	Ground	
	Deg	
	Deix	prox, dist
PATH	Src	±
	End	±

Table 6.4: Features of Place/Path

behind:
$$\begin{bmatrix} \text{PLACE} & \begin{bmatrix} \text{AxPart} & \text{back} \\ \text{LOC} & + \end{bmatrix} \end{bmatrix}$$

house:
$$\begin{bmatrix} \text{PLACE} & \text{Ground} \end{bmatrix}$$

5 meters:
$$\begin{bmatrix} \text{PLACE} & \text{Deg} \end{bmatrix}$$

Figure 6.8: Distribution of features over tokens

- from behind (English)
  Src | Loc & AxPart (2 tokens, 3 features)

- $pic^{h}e$ *se* (Urdu)
  AxPart & Loc | Src (2 tokens, 3 features)

- *pɪc*^{h}-*ũ* (Saraiki)
  AxPart & Loc & Src (1 token, 3 features)

There is a historical explanation for assigning two features to the token *behind*. Actually the *be*-series of adpositions in English comprising prespositions like *behind*, *before*, *beneath* and *between* emerged from a preposition *be* plus some noun, directional particle or some quantifier. Svenonius (2006) has speculated that *be* in these examples served as a Place head, and different lexical elements were recruited to serve as AxParts. A similar situation is with the *a*-series of prepositions (*about*, *above*, *along*, etc) in English.

In line with the proposed selection of head and the LEX-SEM features, the equivalent f-structure representation for the expression '5 meters behind the house' will look

$$\begin{bmatrix} \text{PRED} & \text{`behind}\left\langle\text{house}\right\rangle\text{'} \\ \text{LEX-SEM} & \begin{bmatrix} \text{PLACE} & \begin{bmatrix} \text{AxPart} & \text{back} \\ \text{Loc} & + \\ \text{Deg} & [\;]_1 \\ \text{Ground} & [\;]_2 \end{bmatrix} \end{bmatrix} \\ \text{ADJUNCT} & \begin{bmatrix} \text{PRED} & \text{`meter'} \\ \text{SPEC} & \begin{bmatrix} \text{NUMBER} & 5 \end{bmatrix} \end{bmatrix}^1 \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`house'} \end{bmatrix}_2 \end{bmatrix}$$

Figure 6.9: proposed f-structure for English PP

as in Figure 6.9. This f-structure encodes different notions of spatial expressions in the spirit of the LFG formalism.

Consider now the Urdu spatial expressions in (34), whose f-structures are provided in Figures 6.10 and 6.11.

(34) a. گھر کے ۵ میٹر پیچھے

g^{h}ɑr=ke	5 mɪṭɑr pic^{h}e
house.M.3Sg=Gen.Obl	5 meter behind

‘five meters behind the house’

b. گھر کے ۵ میٹر پیچھے سے

g^{h}ɑr=ke	5 mɪṭɑr pic^{h}e=se
house.M.3Sg=Gen.Obl	5 meter behind=from

‘from five meters behind the house’

The feature ‘Ground’ occurs with the spatial ground irrespective of the fact that it is an OBJ in the English f-structure (Figure 6.9) and an OBL in the Urdu f-structure (Figure 6.10). The explicit morphemes carrying only Loc features in English are basic prepositions, but in Urdu are case markers. Hence, even though the languages differ at the level of c-str, a parallel f-structure representation of spatial expressions is proposed in line with Svenonius’ notions of space, but in the spirit of LFG and the ParGram project.

$$\begin{bmatrix} \text{PRED} & \text{`pic}^{\text{h}}\text{e}\left\langle \text{g}^{\text{h}}\text{ɑr} \right\rangle\text{'} \\ \text{LEX-SEM} & \begin{bmatrix} \text{PLACE} & \begin{bmatrix} \text{AxPart} & \text{back} \\ \text{Loc} & + \\ \text{Deg} & \left[\;\right]_1 \\ \text{Ground} & \left[\;\right]_2 \end{bmatrix} \end{bmatrix} \\ \text{ADJUNCT} & \begin{bmatrix} \text{PRED} & \text{`mɪṭɑr'} \\ \text{SPEC} & \begin{bmatrix} \text{NUMBER} & 5 \end{bmatrix} \end{bmatrix}_1 \\ \text{OBL} & \begin{bmatrix} \text{PRED} & \text{`g}^{\text{h}}\text{ɑr'} \\ \text{CASE} & \text{gen} \end{bmatrix}_2 \end{bmatrix}$$

Figure 6.10: f-structure for Urdu spatial PP in (34a)

$$\begin{bmatrix} \text{PRED} & \text{`pic}^{\text{h}}\text{e}\left\langle \text{g}^{\text{h}}\text{ɑr} \right\rangle\text{'} \\ \text{LEX-SEM} & \begin{bmatrix} \text{PLACE} & \begin{bmatrix} \text{AxPart} & \text{back} \\ \text{Loc} & + \\ \text{Deg} & \left[\;\right]_1 \\ \text{Ground} & \left[\;\right]_2 \end{bmatrix} \\ \text{PATH} & \begin{bmatrix} \text{Src} & + \\ \text{End} & - \end{bmatrix}_3 \end{bmatrix} \\ \text{ADJUNCT} & \begin{bmatrix} \text{PRED} & \text{`mɪṭɑr'} \\ \text{SPEC} & \begin{bmatrix} \text{NUMBER} & 5 \end{bmatrix} \end{bmatrix}_1 \\ \text{OBL} & \begin{bmatrix} \text{PRED} & \text{`g}^{\text{h}}\text{ɑr'} \\ \text{CASE} & \text{gen} \end{bmatrix}_2 \\ \text{CASE} & \text{abl}\ _3 \end{bmatrix}$$

Figure 6.11: f-structure for Urdu spatial PP in (34b)

6.4 Non-spatial adpositions

In addition to the spatial adpositions described and analyzed above, there is a large variety of other adpositions in Urdu. Most of the morphemes used as non-spatial adpostions in Urdu originate from Arabic. In the following subsections, first simple and then complex adpositions in Urdu are discussed.

6.4.1 Simple adpositions

With simple adpositions, I mean here the mono-morphemic adpositions in Urdu. These adpositions license different case marked noun phrases as their complements. Some examples are given in (35).

(35) a. سیلاب کے باعث فصلیں تباہ ہو گئیں.

sɛlab=ke **ba'ɪs** fɑslẽ tɑbah ho gɑĩ
flood=Gen.Obl cause crop.F.3Pl destroy become go-Perf.F.3Pl
'Crops destroyed due to floods'

b. ملکی مفاد کی خاطر غیر مقبول فیصلے کیے.

mʊlki mɑfad=ki **xatɪr** ɣɛr-mɑqbul fɛsle kɪy-e
of-country benefit.M=Gen.F sake.F unpopular decision.3Pl make-Perf.3Pl
'We made unpopular decisions for the benefit of the country.'

c. آج سے / کے بعد

aj=se/ke **bɑ'd**
today=Abl/Gen after
'after today'

d. آج سے / *کے پہلے

aj=se/*ke **pɑhle**
today=Abl/*Gen before
'before today'

The Arabic morphemes *ba'ɪs* 'cause' and *xatɪr* 'sake' are used both as nouns as well as postpostions in Urdu. In contemporary Urdu, however, *xatɪr* is rarely used as a noun and *ba'ɪs* is used as a relational noun only in its singular form as in *ʃaram=ka ba'ɪs* 'cause of embarrassment'. When used as postpositions, the complements are marked with genitive oblique case forms (35a–b) according to the gender these adpositions have as nouns in Urdu. The postposition *ba'd* 'after' can take either an ablative or a genitive marked complement (35c) and the postposition *pahle* 'before' takes only an ablative

complement (35d). The expression *ba'd mẽ* is used as an adverb in the meaning of ‘afterwards’.

The degree or measure phrase that was discussed in case of spatial adpositions can also be involved in temporal adpositional phrases as in (36). In such cases, the reference time can sometimes be omitted.

(36) a. (آج سے/ کے) ایک ہفتہ بعد
(aj=se/ke) ek hɑftɑh **bɑ'd**
today=Abl/Gen one week after
‘a week after (today)’

b. (آج سے/ *کے) ایک ہفتہ پہلے
(aj=se/*ke) ek hɑftɑh **pɑhle**
today=Abl/*Gen one week before
‘a week before (today)’

The Perso-Arabic temporal adpositions *ba'd* ‘after’ and *qabl* ‘before’ used in Urdu sometimes take *az*-prepostional arguments (37).

(37) a. بعد از نماز جمعہ ریلی نکالی گئی.
bɑ'd ɑz nɑmaz-e jʊm'ɑh reli nɪkali gɑ-i
after from prayer-Ez Friday rally hold-Perf.F go-Perf.F.Sg
‘A rally was held after Friday prayer.’

b. ملزم نے قبل از گرفتاری معافی مانگی.
mʊlzɪm=ne **qɑbl** ɑz gɪrɪftari mʊ'afi mãg-i
accused.M=Erg before from arrest.F pardon beg-Perf.F.3Sg
‘The accused begged pardon before arrest.’

(38) a. ندا کا جواب نہ دینا شرم کے باعث ہے.
nɪda=ka jɑvab nɑh de-na ʃɑrɑm=ke
Nida=Gen.M.Sg reply.M.Sg not give-Inf.M.Sg shyness.M=Gen.Obl
ba'ɪs hɛ
because be.Pres.3Sg
‘No reply from Nida is because of shyness.’

b. ندا کا جواب نہ دینا شرم کا باعث ہے۔
nıda=ka jɑvab nɑh de-na ʃɑrɑm=ka ba'ıs hɛ
Nida=Gen.M reply.M not give-Inf.M shyness.M=Gen.M cause be.Pres.3Sg
'No reply from Nida is the cause of embarrassment.'

c. ندا کا جواب نہ دینا باعث شرم ہے۔
nıda=ka jɑvab nɑh de-na ba'ıs-e ʃɑrɑm hɛ
Nida=Gen.M reply.M not give-Inf.M casue-Ez embarrassment.M be.Pres.3Sg
*'No reply from Nida is because of shyness.'
'No reply from Nida is the cause of embarrassment.'

Some adpositions taking genitive marked complements also allow complements in the *ezafe* construction. For example, *ba'ıs* as an adposition allows a genitive marked complement (38a), but does not allow the complement in the *ezafe* construction (38c). So, *ba'ıs-e ʃaram* would be ungrammatical in the meaning of 'because of embarrassment'. However, the word *ba'ıs* as a noun takes either a genitive complement or its complement is in the *ezafe* construction: *ʃaram=ka ba'ıs* 'cause of embarrassment' or *ba'ıs-e ʃaram* 'cause of embarrassment'. An example of an adposition either taking a genitive complement or allowing its complement in the *ezafe* construction is given in (39).

(39) a. میچ کا نتیجہ توقع کے خلاف ہے۔
mɛc=ka nɑtijɑh tɑvaqqʊ'=ke **xılaf** hɛ
match.M.Sg=Gen.M.Sg result.M expectation.F=Gen.Obl against be.Pres.3Sg
'The result of the match is unexpected.'

b. میچ کا نتیجہ خلاف توقع ہے۔
mɛc=ka nɑtijɑh **xılaf**-e tɑvaqqʊ' hɛ
match.M=Gen.M result.M against-Ez expectation.F be.Pres.3Sg
'The result of the match is unexpected.'

Another example of an adposition that allows both types of complements is *dɔran* 'during'. The phrases *gʊftgu=ke dɔran* and *dɔran-e gʊftgu* both mean 'during conversation'. In Urdu there is only one non-spatial preposition *hasb* 'as per/according to'

which requires an *ezafe* construction to make a phrase with its complement. To show that this pattern is productive in Urdu, examples from an Urdu corpus are listed in Table 6.5.

hasb-e ɪradah 'as per determination'	*hasb-e ʊsul* 'as per rule'
hasb-e taqaza 'as per requirement'	*hasb-e tɔfiq* 'as per capability'
hasb-e tavaqqʊ 'as per expectation'	*hasb-e hal* 'as per time'
hasb-e hɛsɪyyat 'as per status'	*hasb-e dastur* 'as per standard'
hasb-e zaɪqah 'as per taste'	*hasb-e rɪvaj* 'as per fashion'
hasb-e sabɪq 'as per previous'	*hasb-e rɪvayat* 'as per tradition'
hasb-e adat 'as per habit'	*hasb-e zabɪtah* 'as per precept'
hasb-e zarurat 'as per need'	*hasb-e tariqah* 'as per method'
hasb-e qanun 'as per law'	*hasb-e qaɪdah* 'as per regulation'
hasb-e qavaɪd 'as per regulations'	*hasb-e mazi* 'as per past'
hasb-e ma'mul 'as per usual'	*hasb-e manʃa* 'as per wish'
hasb-e mansubah 'as per plan'	*hasb-e mɔqa* 'as per opportunity'
hasb-e nɪyyat 'as per intention'	*hasb-e va'dah* 'as per promise'
hasb-e ɪradah o mansubah 'as per determination and plan'	
hasb-e ʊsul o qavaɪd 'as per rules and regulations'	
hasb-e marzi o manʃa 'as per will and wish'	

Table 6.5: Use of the adposition *hasb* 'as per' in Urdu

The adposition *bɪn/bɪna* 'without' is one of some adpositions in Urdu which can take either a nominative complement as in *tʊm bɪn* 'without you' or a genitive marked complement as in *tʊmhare bɪn* 'without you'. This adposition, like another adposition *baɣɛr* 'without', can also take an oblique form of a perfective participle as its complement, as in *bɪn dek*^{h}*-e* 'without seeing' or *dek*^{h}*-e baɣɛr* 'without seeing'.

6.4.1.1 Attachment of adpositional phrases

Some adpositional phrases can modify both nouns and verbs and others can only modify verbs. The use of the adposition *mʊt'alɪq* 'about' is illustrated in (40).

(40) a. عدالت نے ملزم کی جائیداد سے متعلق پوچھا.

ɑdalɑt=ne [mʊlzɪm=ki jaɪdad=se **mʊt'ɑlɪq**] pucʰ-a
court.F=Erg accused.M=Gen.F property.F=Abl about ask-Perf.M.3Sg
'The court asked about the property of the accused.'

b. عوام نے حکومت کی روزگار سے متعلق پالیسی کو پسند کیا.

ɑvam=ne [hɑkumɑt=ki rozgar=se **mʊt'ɑlɪq**
public.F=Erg government.F=Gen.F employment.M=Abl about
palisi]=ko pɑsɑ̃d kɪy-a
policy.F=Acc like do-Perf.M.3Sg
'People praised the employment policy of the government.'

c. بنیادی انسانی حقوق سے متعلقہ قوانین

bʊnyadi ɪnsani hɑquq=se mʊt'ɑlɪqɑh qɑvanin
basic of-human right.M.Pl=Abl related law.M.pl
'Laws related to the basic human rights'

In (40a) the adpositional phrase *mʊlzɪm=ki jaɪdad=se mʊt'ɑlɪq* 'about the property of the accused' is in fact modifying the verb or in a sense the whole sentential clause. In (40b) the adpositional phrase *rozgar=se mʊt'ɑlɪq* modifies the noun *palisi* 'policy'. There is another equivalent word *mʊt'ɑlɪqɑh* 'related' in Urdu which is used as an adjective (40c). The use of another adposition *xɪlaf*, which is only attached at the clause level, and an adjective *mʊxalɪf* is exemplified in (41)–(42).

(41) a. حکومت کے مخالف لوگوں نے حکومت کے خلاف جلوس نکالا.

hɑkumɑt=ke **mʊxalɪf** logõ=ne hɑkumɑt=ke
government.F=Gen.3Pl opposing people.3Pl=Erg government.F=Gen.Obl
xɪlaf jʊlus nɪkal-a
against procession.M conduct-Perf.M
'People opposed to the government conducted a procession against the government.'

b. حکومت کی مخالف جماعت نے حکومت کے خلاف ریلی نکالی۔

hɑkumɑt=ki **mʊxalıf** jɑma'ɑt=ne hɑkumɑt=ke
government.F=Gen.F.3Sg opposing party.F=Erg government.F=Gen.Obl

xılaf reli nıkal-i
against rally.F conduct-Perf.F

'The party opposed to the government conducted a rally against the government.'

(42) a. عوام کی اکثریت جنگ کی مخالف ہے۔

ɑvam=ki ɑksɑrıyɑt jɑ̃g=ki **mʊxalıf** hɛ
public.F=Gen.F majority.F war.F=Gen.F opposed.F be.Pres
'The majority of people are opposed to war.'

b. عوام کی اکثریت جنگ کے خلاف ہے۔

ɑvam=ki ɑksɑrıyɑt jɑ̃g=ke **xılaf** hɛ
public.F=Gen.F majority.F war.F=Gen.Obl against be.Pres
'The majority of people are against war.'

In (41), the complement of the adposition is marked by an oblique genitive marker, however, the argument of the adjective is marked genitive according to the gender and number of the modified head-noun. The argument of the adjective in (41a) is marked for 3Pl genitive marker due to plural head noun *logõ* 'people' and in (41b) is marked for F.3Sg due to the feminine head noun *jama'at* 'party'.

The adjective *mʊxalıf* 'opposed' is used predicatively in (42a). In this case also, the complement of the adjective is marked genitive according to the number and the gender of the subject *aksarıyat* 'majority'. In (42b), however, the complement of the adposition is marked by an oblique genitive marker.

It should be noted, as already said, that complements of some adpositions are marked for feminine genitive marker in the case adposition is originally a feminine noun, however, when the complement is postposed the default oblique masculine form is used (Khan, 2005), for example *ali=ki mısl* 'like Ali' and *mısl ali=ke* 'like Ali'.

6.4.1.2 *ɪs lɪye* ‘therefore’ and *ɪs tarah* ‘thus’

The adpositions *lɪye* ‘for’ and *tarah* ‘like’ take genitive marked complements in Urdu. Their canonical use is illustrated in (43).

(43) a. .علی ندا/ اس کے لیے پھول لایا

ɑli [nɪda/ʊs]=ke lɪye p^{h}ul la-ya
Ali.M Nida.F/3Sg=Gen for flower.M.3Sg bring-Perf.M.3Sg
‘Ali brought a flower for Nida/her.’

b. .علی/ اس کی طرح ندا بھی ذہین ہے

[ɑli/ʊs]=ki tɑrɑh nɪda b^{h}i zɑhin hɛ
Ali.M/3Sg=Gen.F like Nida.F also intelligent be.Pres.3Sg
‘Nida is also intelligent like Ali/him.’

The phrase *ɪs/ʊs ke lɪye* means ‘for him/her/it’ and *ɪs/ʊs ki tarah* means ‘like him/her/it’. So, the meanings of these phrases are constructed syntactically. However, the word combination *ɪs lɪye* is also used in Urdu in the lexical meaning of ‘therefore/hence’ and *ɪs tarah* is used in the lexical meaning of ‘thus/so’. Two other expressions *kɪs lɪye* and *kɪs tarah* are also used with the meanings ‘why’ and ‘how’ respectively. Each pair of these lexical expressions show a similar syntactic behavior, as shown in (44)–(47).

(44) a. .ندا بیمار ہے، اس لیے کالج نہیں آئی

nɪda bimar hɛ ɪs lɪye kalɪj nɑhĩ a-i
Nida.F ill be.Pres.3Sg therefore college.M not come-Perf.F.3Sg
‘Nida is ill, therefore, she did not came to college.’

b. .علی نے محنت کی، اس طرح کامیاب ہو گیا

ɑli=ne mɪhnɑt ki ɪs tɑrɑh kamyab ho gɑ-ya
Ali.M=Erg hardwork.F do-Perf.F.Sg thus successful become go-Perf.M.3Sg
‘Ali did hard work, so, succeeded.’

(45) a. ندا یہاں کس لیے آئی؟

nɪda yɑhɑ̃ kɪs lɪye a-i
Nida.F here why come-Perf.F.3Sg
'Why did Nida come here?'

b. ندا یہاں کس طرح آئی؟

nɪda yɑhɑ̃ kɪs tɑrɑh a-i
Nida.F here how come-Perf.F.3Sg
'How did Nida come here?'

(46) a. ندا یہاں علی سے ملنے کے لیے آئی.

nɪda yɑhɑ̃ ɑli=se mɪl-ne=ke lɪye a-i
Nida.F here Ali.M=Com meet-Inf.Obl=Gen.Obl for come-Perf.F.3Sg
'Nida came here to meet Ali.'

b. ندا یہاں چوروں کی طرح آئی.

nɪda yɑhɑ̃ corõ=ki tɑrɑh a-i
Nida.F here thief.M.3Pl=Gen.F like come-Perf.F.3Sg
'Nida came here like a thief.'

(47) a. ندا یہاں اس لیے آئی کہ علی سے ملے.

nɪda yɑhɑ̃ ɪs lɪye a-i kɪh ɑli=se mɪl-e
Nida.F here hence come-Perf.F.3Sg that Ali.M=Com meet-Subjn
'Nida came here, so that she meet Ali.'

b. ندا یہاں اس طرح آئی کہ کسی کو پتا نہیں چلا.

nɪda yɑhɑ̃ ɪs tɑrɑh a-i kɪh kɪsi=ko pɑta nɑhĩ cɑl-a
Nida.F here so come-Perf.F that any=Dat knowledge not walk-Perf.M
'Nida came here in a way that no one noticed.'

There is still another multiword expression of this series *jɪs tarah*, which has the meaning of 'as' as in *jɪs tarah ap cahẽ* 'as you wish'. Its counterpart expression with

the *lıye* adposition is *jıs lıye*. Note that *ıs tarah*, *kıs tarah* and *jıs tarah* are in fact reduced forms of *ıs tarah se* 'by this way', *kıs tarah se* 'by which way' and *jıs tarah se* 'by which (ever) way', respectively.

6.4.2 Complex adpositions

In many languages, the basic adpositions together with nouns seem to act as predicators and take some other elements as complements. Such predicators in linguistic literature (Kurzon & Adler, 2008; Quirk & Mulholland, 1964) are usually termed complex adpositions or compound adpositions or phrasal prepositions. Some examples of these are listed in the following:

- French: *en face de*, *en dépit de*, *au milieu de*
- Spanish: *al lado de*, *en casa de*
- Swedish: *i början av*, *med hjälp av*, *i stället för*
- English: *in view of*, *in spite of*, *by dint of*
- Polish: *bez wzgledu na*, *w zwiazku z*, *z uwagi na*
- German: *an Hand von*, *mit Hilfe von*, *in Bezug auf*

(adapted from Trawinski 2003)

Table 6.6 shows examples of complex prepositions from Persian, many of which are also used in Urdu. In these complex prepositions a noun is added to a basic preposition and the combination then licenses a complement, usually NP, mostly in the *ezafe* construction. For other types of complex presspositions in Persian, see Chime (2006).

The use of the first two complex prepositions in Table 6.6 are illustrated in (48). These complex prepositions are hardly found in Urdu news corpus, however, they are found in poetry and in the religious literature written in Urdu.

Preposition	Complex prepositions
bar 'on'	*bar sabil* 'in the way of/as', *(bar) sar* 'on', *bar aks* 'opposite to' *bar asas* 'based on', *(bar) xılaf* 'against', *(bar) tıbq* 'according to'
dar 'in'	*(dar) mahzar* 'in presence of/to', *(dar) xıdmat* 'in service of/to' *dar murıd* 'in case/respecting', *dar barıh* 'about'
az 'from'	*az jıht* 'with regard to/for', *az rah* 'by the way of' *az ru* 'with respect to', *az janıb* 'from the side of/from'
bıh/bah 'to'	*(bıh) jıht* 'for the sake of/for', *ba-tɔr* 'by way of" *ba-qɔl* 'according to', *ba-ja* 'in place of"
ba 'with'	*ba vajud* 'with existence of/in spite of'

Table 6.6: Persian complex prepositions

(48) a. ہے مجھ کو تجھ سے تذکرہ غیر کا گلہ

ہرچند برسبیل شکایت ہی کیوں نہ ہو

hɛ	mʊj^h=ko	tʊj^h=se	tazkarah-e	ɣɛr=ka	gılah
be.Pres	1Sg=Acc	2Sg=Abl	mention-Ez	rival=Gen.M	remonstrance.M

har-cãd	**bar-sabil-e**	**ʃıkayat**	hi	kıyɔ̃	nah	ho
even	on-way-Ez	complaint.F	Emp	why	not	be.Subjn

'I remonstrate with you for the mention of the rival;
even if his complaint is made in this mention.' (Asadullah Khan Ghalib)

b. ہونٹوں پہ کبھی ان کے میرا نام ہی آئے

آئے تو سہی برسر الزام ہی آئے

hõṭõ=pıh	kab^{h}i	ʊn=ke	mera	nam	hi	ae
lip.3Pl=Loc.on	ever	3Pl=Gen.Obl	my	name.M	Emp	come.Subjn

ae	to	sahi	**bar-sar-e**	**ılzam**	hi	ae
come.Subjn	Emp	Emp	on-head-Ez	blame	Emp	come.Subjn

'May my name come on her/his lips;
it should come, even if it come as a start of blame.' (Ada Jafari)

In these examples we see that complements of complex prespositions occur in the *ezafe* construction. Some of the Persian complex prepositions can take oblique genitive

marked complements in Urdu. For example, *bar aks* 'opposite to' takes a genitive marked complement as in *ma'mul=ke bar aks* 'against usual'. Another example is given in (49b). The genitive complements of complex adpositions can also be postposed (49c).

(49) a. بقول علی ندا بیمار ہے.
bɑ-qɔl(-e) ɑli nɪda bimar hɛ
to-saying.M(-Ez) Ali Nida ill be.Pres.3Sg
'According to Ali, Nida is ill.'

b. علی کے بقول ندا بیمار ہے.
ɑli=ke bɑ-qɔl nɪda bimar hɛ
Ali.M=Gen.Obl.M to-saying.M Nida ill be.Pres.3Sg
'According to Ali, Nida is ill.'

c. بقول علی کے ندا بیمار ہے.
bɑ-qɔl ɑli=ke nɪda bimar hɛ
to-saying.M Ali.M=Gen.Obl.M Nida ill be.Pres.3Sg
'According to Ali, Nida is ill.'

The form of the oblique genitive marker which marks the complement of a complex adposition is according to the gender and number of the noun, which is part of complex adposition. The noun *qɔl* 'saying' in (49b–c) is masculine, so the masculine oblique form *ke* is used to mark the complement *ɑli* 'Ali'. Another example of a prepositional phrase is *ɑli=ki bajae* or *bajae ɑli=ke* 'instead of Ali'; here the complement at its canonical position is marked by feminine form of genitive marker because the noun in the complex adposition is feminine.[2]

There are some Persian complex prepositions which in their original form do not take genitive marked complements. Their equivalent Urdu forms can only take genitive complements. Such Persian complex prepositions and their Urdu equivalents are given in Table 6.7.

The Persian complex adposition *ba-qɔl* 'according to' does not have a literal equivalent in Urdu. Like *ba-qɔl* 'according to', the complex lexical sign *ba-tɔr* 'by way of' can have its complement in an *ezafe* construction or be genitively marked. Its Urdu equivalent is *tɔr par* 'by way of'. The peculiarity with this sign is that the noun part of it can be modified by an adjective both in Persian and Urdu and in that case it does not take its nominal complement, for example, *ba-tɔr-e xas* 'as a special case' in Persian

[2]The Persian complex preposition *baja* is lexicalized as *bajae* in Urdu. That, is the *ezafe* clitic has become the part of the adposition and remains there even if its complement is genitive marked.

Persian complex preposition	Urdu equivalent
bar asas 'based on'	*asas par* 'based on'
dar xıdmat 'in service of/to'	*xıdmat mẽ* 'in service of/to'
dar barıh 'in behlaf/about'	*bare (mẽ)* 'about'
az ru 'with respect to'	*ru se* 'with respect to'
az janıb 'from the side of/from'	*janıb se* 'from the side of/from'

Table 6.7: Persian complex prepositions and their Urdu equivalents

and *xasusi tɔr par* 'as a special case' in Urdu. However this property is generally not found with complex adpositions.

So, many complex adpositions in Urdu actually have been borrowed from Persian either in the form as they are in Persian, or the basic adposition part of them is replaced by an Urdu case clitic. To show that these combinations are really complex, I now present the syntactic distribution of nouns in them.

6.4.2.1 Syntactic distribution of nouns in complex adpositions

The nouns in complex adpositions display a different syntactic distribution as compared to the syntactic distribution of nouns elsewhere. Chime (2006) has discussed such differences for Persian that apply to Urdu as well. Consider a simple phrase and a complex adpositional phrase both in Persian (50-a1,b1) and Urdu (50-a2,b2).

(50) a1. بر دیوار خانه
bɑr divar-e xanıh
on wall-Ez house
'on the wall of the house' (Persian)

a2. گھر کی دیوار پر
g^hɑr=ki divar **pɑr**
house=Gen wall on
'on the wall of the house'

b1. بر اساس مطالعات
bɑr ɑsas-e mʊtalı'at
on base-Ez research.Pl
'based on research' (Persian)

b2. تحقیقات کی اساس پر
tɑhqiqat=ki **ɑsas pɑr**
research.Pl=Gen base on
'based on research'

The nominal part of complex adpositions cannot be modified by demonstratives as is shown in (51).

(51) a1. بر (این) دیوار خانه
bɑr (in) divar-e xanıh
on (this) wall-Ez house
'on (this) wall of the house' (Persian)

a2. گھر کی (اس) دیوار پر
gʰɑr=ki (is) divar **pɑr**
house=Gen (this) wall on
'on (this/one/every) wall of the house'

b1. بر (*این) اساس مطالعات
bɑr (*in) **ɑsas**-e mʊtalı'at
on (*this) base-Ez research.Pl
'based on research' (Persian)

b2. تحقیقات کی (*اس) اساس پر
tɑhqiqat=ki (*is) **ɑsas pɑr**
research.Pl=Gen (*this) base on
'based on research'

Nor can nouns of complex adpositions take an indefinite article (in case of Persian) or quantifiers (52).

(52) a1. بر (یک/ ہر) دیوار خانه
bɑr (yık/hɑr) divar-e xanıh
on (one/every) wall-Ez house
'on (one/every) wall of the house' (Persian)

b1. گھر کی (ایک/ ہر) دیوار پر
gʰɑr=ki (ek/hɑr) divar **pɑr**
house=Gen (one/every) wall on
'on (one/every) wall of the house'

b1. بر (*یک/ *ہر) اساس مطالعات
bɑr (*yık/*hɑr) **ɑsas**-e mʊtalı'at
on (*one/*every) base-Ez research.Pl
'based on research' (Persian)

b2. تحقیقات کی (*ایک/ *ہر) اساس پر
tɑhqiqat=ki (*ek/*hɑr) **ɑsas** **pɑr**
research.Pl=Gen (*one/*every) base on
'based on research'

Neither question words nor words of exclamation can modify nouns of complex adpositions (53).

(53) a1. بر (چه) دیوار خانه
bɑr (cıh) divar-e xanıh
on (what) wall-Ez house
'on (what) wall of the house' (Persian)

b1. گھر کی (کس) دیوار پر
gʰɑr=ki (kis) divar **pɑr**
house=Gen (what) wall on
'on (what) wall of the house'

b1. بر (*چه) اساس مطالعات
bɑr (*cıh) **ɑsas**-e mʊtalı'at
on (*what) base-Ez research.Pl
'based on research' (Persian)

b2. تحقیقات کی (*کس) اساس پر
tɑhqiqat=ki (*kis) **ɑsas pɑr**
research.Pl=Gen (*what) base on
'based on research'

Adjectival modification of nouns in complex adpositions is also not possible (54).

(54) a1. بر دیوار (بزرگ) خانه
bɑr divar-e (bʊzʊrg-e) xanıh
on wall-Ez (big-Ez) house
'on the (big) wall of the house'
(Persian)

a2. گھر کی (بڑی) دیوار پر
g^hɑr=ki (bɑṛi) divar **pɑr**
house=Gen (what) wall on
'on the (big) wall of the house'

b1. بر اساس (*بزرگ) مطالعات
bɑr asas-e (*bʊzʊrg-e) mʊtalı'at
on base-Ez (*big) research.Pl
'based on research' (Persian)

b2. تحقیقات کی (*بڑی) اساس پر
tɑhqiqat=ki (*bɑṛi) **asas pɑr**
research.Pl=Gen (*big) base on
'based on research'

The noun of complex adpositions cannot be replaced by a plural form (55).

(55) a1. بر دیوار (های) خانه
bɑr divar-(ha-y)e xanıh
on wall-Pl-Ez house
'on wall(s) of the house' (Persian)

a2. گھر کی دیوار(وں) پر
g^hɑr=ki divar(õ) **pɑr**
house=Gen wall(Pl) on
'on wall(s) of the house'

b1. بر اساس (*های) مطالعات
bɑr asas-(*ha-y)e mʊtalı'at
on base-*Pl-Ez research.Pl
'based on research' (Persian)

b2. تحقیقات کی اساس (*وں) پر
tɑhqiqat=ki **asas**(*õ) **pɑr**
research.Pl=Gen base(*Pl) on
'based on research'

Coordination of another noun with the noun in the complex adposition is also not possible (56b).

(56) a1. بر (بام و) دیوار خانه
bɑr (bam o) divar-e xanıh
on (roof and) wall-Ez house
'on (roof and) wall of the house'
(Persian)

a2. گھر کی (چھت اور) دیوار پر
g^hɑr=ki (c^hɑt ɔr) divar **pɑr**
house=Gen (roof and) wall on
'on (roof and) wall of the house'

b1. بر (*بنا و) اساس مطالعات
bɑr (*bına o) **asas**-e mʊtalı'at
on (*basis and) base-Ez research.Pl
'based on research' (Persian)

b2. تحقیقات کی (*بنیاد اور) اساس پر
tɑhqiqat=ki (*bʊnyad ɔr)
research.Pl=Gen (*basis and)
asas pɑr
base on
'based on research'

These syntactic differences support an analysis in which the two lexical categories in Persian *bar* 'on' and *asas* 'basis' can be assumed to project to a complex adposition *bar asas* which heads the entire phrase. The lexically selected complement of the relational noun in such phrases is syntactically realized as the object of the complex adposition. Likewise, in the case of similar Urdu constructions, the relational noun can also be assumed to be part of the complex postposition.

6.4.2.2 *zer*+N complex adpositions

There are some constructions in which the preposition *zer* 'under' and a noun seem to act like a complex adposition. Complex adpositions of this type are more common in Urdu than in Persian. One of such complex adpositions is *zer-e nazar*[3] 'under supervision/guidance/vigilance of' which is found in Persian (57a) as well as in Urdu (57b). It should be noted that the phrase *zer-e nazar* is also used as an attributive adjective in the meaning of 'what is being sighted' (58).

(57) a. ندا زیر نظر استاد این مقاله را ادامه داد.

nıda	**zer-e**	**nazar**-e	ʊstad	in	maqalıh	ra
Nida.F.3Sg	under-Ez	supervision-Ez	professor.3Sg	this	article.3Sg	Acc
ıdamıh	dad					
completion	give.3Sg.Past					

'Nida completed this article under the supervision of a professor.' (Persian)

b. ندا نے پروفیسر کی/ کے زیر نظر یہ مقالہ مکمل کیا.

nıda=ne	parofɛsar=ki/ke	**zer-e**	**nazar**	yıh	maqalah
Nida.F=Erg	professor.3Sg=Gen	under-Ez	supervision.F	this	article.M.3Sg
mʊkammal	kı-ya				
complete	do-Perf.M.3Sg				

'Nida completed this article under the supervision of a professor.'

c. ندا نے پروفیسر کی نظر کے زیر یہ مقالہ مکمل کیا.

* nıda=ne	parofɛsar=ki	**nazar=ke**	**zer**	yıh	maqalah
Nida.F=Erg	professor.3Sg=Gen	sight.F=Gen	under	this	article.M.3Sg
mʊkammal	kı-ya				
complete	do-Perf.M.3Sg				

'Nida completed this article under the supervision of a professor.'

[3] The Persian preposition *zer* can occur with or without *ezafe* (Pantcheva, 2006). In the LFG implementation which will come later, I have considered *zer* and *zer-e* as two forms of the same preposition.

(58) زیر نظر کتاب کا مسودہ ۱۹۰۰ میں لکھا گیا.

zer-e	**nɑzɑr**	kɪtab=ka	mʊsɑvvɑdɑh	1900=mẽ
under-Ez	sight.F.3Sg	book.F.3Sg=Gen.M.3Sg	draft.M.3Sg	1900=Loc.in

lɪkʰ-a	gɑ-ya
write-Perf.M.3Sg	go-Perf.M.3Sg

'The original draft of the book, which is being examined, was written in 1900.'

All the syntactic differences discussed for the noun part of complex adpositions in the previous section are also applicable to the noun *nazar* 'supervision/consideration' in *zer-e nazar* 'under supervision/guidance/vigilance of'. I provide some more arguments in favor of the complex adposition analysis of *zer-e nazar* in Urdu.

The complement of the complex adposition *zer-e nazar* in Urdu (57b) is marked by the genitive case, the form of which is either feminine according to the gender of the noun *nazar* 'sight' or the default masculine oblique. The Persian prepositional phrase *zer-e nazar-e ʊstad* 'under the supervision of the professor' cannot be translated as *ʊstad ki nazar ke zer* (57c) according to the pattern by which the Persian noun phrase in (59a) is translated in Urdu in (59b).

(59) a. گنبد کلیسای شهر خیلی بلند است.

gʊmbɑd-e	kɑlisa-ye	ʃɑhr	xɛli	bʊlɑ̃d	ɑst
dome.3Sg-Ez	church.3Sg-Ez	city.3Sg	very	high	be.Pres.3Sg

'The dome of the church of the city is very high.' (Persian)

b. شہر کے کلیسے کا گنبد کافی اونچا ہے.

ʃɑhr=ke	kɑlise=ka	gʊmbɑd	kafi	ũca	hɛ
city.M=Gen.Obl	church.M=Gen.M.Sg	dome.M.3Sg	very	high	be.Pres.3Sg

'The dome of the church of the city is very high.'

That is the two nouns *nazar* 'supervision' and *ʊstad* which are added to the basic preposition *zer* in an in the *ezafe* construction cannot both be replaced by genitive marked nouns in Urdu in contrast with the other noun phrases. Only the second noun can be replaced by a genitive marked noun in Urdu. To provide one more contrast, I

give some more examples of Persian noun phrases with two modifier nouns in the *ezafe* construction and their equivalent translations in Urdu in (60)–(61).

(60) a1. ذوق شاعری اقبال
zɔq-e ʃa'ɪri-**ye** **ɪqbal**
verve-Ez poetry-**Ez** **Iqbal**
'Iqbal's verve for poetry' (Persian)

a2. اقبال کا ذوق شاعری
ɪqbal=ka zɔq-e ʃa'ɪri
Iqbal=Gen.M verve.M-Ez poetry.F
'Iqbal's verve of poetry.'

b1. طرز حکمرانی پہلوی
tɑrz-e hʊkmɑrani-**ye** **pɑhlvi**
way-Ez rule-**Ez** **Pahlvi**
'The way of rule of Pahlvi' (Persian)

b2. پہلوی کا طرز حکمرانی
pɑhlvi=ka tɑrz-e hʊkmɑrani
Pahlvi=Gen.M.Sg way.M.Sg-Ez rule.F
'The way of rule of Pahlvi'

c1. شجر سایہ دار جنت
ʃɑjr-e sayɑh-dar-**e** **jɑnnɑt**
tree-Ez shady-**Ez** **heaven**
'The heaven's shady tree' (Persian)

c2. جنت کا شجر سایہ دار
jɑnnɑt=ka ʃɑjr-e sayɑhdar
heaven=Gen.M tree.M-Ez shady
'The heaven's shady tree'

(61) a1. لباس پسر علی
lɪbas-e pɪsɑr-**e** **ɑli**
suit-Ez son-**Ez** **Ali**
'The suit of Ali's son' (Persian)

a2. علی کا لباس پسر
* **ɑli=ka** lɪbas-e pɪsɑr
Ali=Gen.M suit.M-Ez son.M
'The suit of Ali's son'

b1. خوشبوی زلف یار
xʊʃbu-ye zʊlf-**e** **yar**
fragrance-Ez hair-**Ez** **beloved**
'The fragrance of the beloved's hair' (Persian)

b2. یار کی خوشبوئے زلف
* **yar=ki** xʊʃbu-ye zʊlf
beloved=Gen.F fragrance.F-Ez hair.F
'The fragrance of the beloved's hair'

c1. درد سر دوست
dɑrd-e sɑr-**e** **dost**
ache-Ez head-**Ez** **friend**
'Friend's headache' (Persian)

c2. دوست کا درد سر
dost=ka dɑrd-e sɑr
friend.M=Gen.M ache.M-Ez head.M
'Friend's headache'

In each of the Persian noun phrases in (60)–(61) there are two modifiers. There is one difference between the noun phrases in (60) and those in (61). In (60) two modifiers modify a single head noun, however, in (61), the first modifier modifies the head noun and the second modifier modifies the first modifier.

In (60) we see that the second nominal modifier of a head noun in Persian examples is replaced by a genitive modifier in Urdu. In (61), however, the second modifier (which actually is the modifier of the modifier of the head noun) cannot be replaced by a genitive modifier in Urdu noun phrases. This is due to the fact that modifiers are always contiguous to the head nouns in Urdu. The second nominal modifier in (60-c1) is replaced by the respecting genitive modifier in Urdu (60-c2) because in Urdu *dard-e sar* 'headache' behaves as a compound word like *vazir-e azam* 'prime-minster'. One could alternatively assume that the two modifiers in (60) are modifying the same head noun *dard* 'pain', however, for me, this claim is not justifiable semantically.

Now, we see how things are with prepositional phrases. Three adpositional phrases are shown in (62). In all three instances of the Persian prepositional phrases the last modifier in fact modifies the object of the preposition *zer* 'under'.

(62) a1. زیر قالین مسجد
zer-e qalin-**e** **mɑsjɪd**
under-Ez carpet-**Ez** **mosque**
'under the carpet of the mosque' (Persian)

a2. مسجد کی/کے زیر قالین
* **mɑsjɪd=ki/ke** zer-e qalin
mosque.F=Gen under-Ez carpet.F
'under the carpet of the mosque'

b1. زیر پای ندا
zer-e pa-**ye** **nɪda**
under-Ez foot-**Ez** **Nida**
'under the foot of Nida' (Persian)

b2. ندا کے زیر پا
* **nɪda=ke** zer-e pa
Nida=Gen.M under-Ez foot.M
'under the foot of Nida'

c1. زیر نظر علی
zer-e nɑzɑr-**e** **ɑli**
under-Ez supervision-**Ez** **Ali**
'under the supervision of Ali' (Persian)

c2. علی کی/کے زیر نظر
ɑli=ki/ke zer-e nɑzɑr
Ali=Gen under-Ez supervision-Ez
'under the supervision of Ali'

In the first two cases, other modifiers can intervene between the object and the existing modifier, for example, *qaʃãg* 'beautiful' can intervene between *qalin* 'carpet' and *masjɪd* 'mosque' to make a prepositional phrase of the meaning 'under the beautiful carpet of the mosque'. So, in the first two cases *zer* acts as a simple preposition and we see that the modifier of the object of this simple preposition cannot be replaced by respecting a genitive complement in Urdu.

In the third example of a Persian prepositional phrase in (62-c1), however, no other modifier can intervene between *nazar* 'supervision' and *ali* 'Ali'. So, in the third case *zer-e nazar* 'under supervision' acts as a compound or complex adposition, the object of which can be replaced by genitive marked object in Urdu. The genitive marker on the

modifier of a compound noun shows agreement with the head noun (61-c1) , however, the genitive marker on the object of a complex adposition can be in masculine oblique form even if the noun in the complex adposition is feminine (62-c1).

(ke)	*zer-e ɑsɑr* 'under influence (of)'
	zer-e ɪstɪ'mal 'used (by)'
	zer-e ɪntɪzam 'under management (of)'
	zer-e ɪhtɪmam 'under management (of)'
	zer-e bar 'under pressure (of)'
	zer-e tasallʊt 'under occupation (of)'
	zer-e ɪlaj 'under curing (of)/cured (by)'
	zer-e ʊnvan 'under title (of)'
	zer-e sayɑh 'under auspices (of)'
	zer-e ɣɔr 'under consideration (of)'
	zer-e qɑbzah 'under occupation (of)'
	zer-e kanṭrol 'under control (of)'
	zer-e mʊtalɪ'ah 'under reading (of)'
	zer-e mʊaɪnah 'under examination (of)'
	zer-e nɑgin 'under rule (of)'
(ki/ke)	*zer-e ɪdarat* 'under editorship (of)'
	zer-e tarbɪyat 'under training (of)'
	zer-e taftiʃ 'under investigation (of)'
	zer-e tavɑjjʊh 'under attention (of)'
	zer-e harasat 'under arrest (of)'
	zer-e hʊkmarani 'under rule (of)'
	zer-e sarpɑrasti 'under mentorship (of)'
	zer-e sama'at 'under hearing (of)'
	zer-e sadarat 'under chair (of)'
	zer-e ʃafqat 'under kindness (of)'
	zer-e qɪyadat 'under leadership (of)'
	zer-e kaʃt 'under cultivation (of)'
	zer-e kaman 'under command (of)'
	zer-e nazarart 'under supervision (of)'
	zer-e nazar 'under supervision (of)'
	zer-e nɪgrani 'under vigilance (of)'
	zer-e hɪdayat 'under guidance (of)'

Table 6.8: *zer-e+N* complex constructions in Urdu

A list of *zer+N* complex adpositions which take genitive marked complements in Urdu is given in Table 6.8. The complex adpostion *zer-e ʊnvan* 'under title (of)' has another alternate construction in Urdu as *ʊnvan ke taht* 'under title (of)'. Except

for this complex adposition, complements of all other complex constructions in the Table 6.8 are somehow animate in nature.

6.4.2.3 *peʃ-e nazar*

The complex construction *peʃ-e nazar* 'in view (of)' is the short form of the Persian expression *dar peʃ-e nazar*. In Urdu it is typical in that it can take two types of arguments, either experiencer or theme.

(63) a. عوام کی / کے پیش نظر جنگ کا خطرہ ہے۔

ɑvam=ki/ke	peʃ-e	nɑzɑr	jɑ̃g=ka	xɑtrɑh	hɛ
public.F=Gen	before	sight.F	war.F=Gen.M.Sg	danger.M.Sg	be.Pres.3Sg

'The danger of war is in sight of the public.'

b. جنگ کے خطرے کے پیش نظر عوام پرشان ہے۔

jɑ̃g=ke	xɑtre=ke	peʃ-e	nɑzɑr	ɑvam	pɑreʃan	hɛ
war.F=Gen.M	danger.M=Gen.Obl	before	sight.F	public.F	worried	be.Pres.3Sg

'The public is worried in light of war.'

The genitive marked argument of *peʃ-e nazar* 'in view (of)' in (63a) is the experiencer *avam* 'public' and in (63b) it is the theme *jɑ̃g=ka xatrah* 'danger of war'. A similar complex construction *madd-e nazar* 'in view (of)' usually takes a theme argument as its complement.

In this section different types of complex adpositions in Urdu have been analyzed. It was shown that the noun part of complex adpositions shows different syntactic distribution in complex adpositions. The *zer+N* complex adpositions were contrasted with the similar Persian noun phrases and their equivalents in Urdu.

6.5 Modeling complex adpositions in LFG

There are many possible ways of dealing with the syntax of the complex predicators discussed in the previous sections. One way not to treat them as complex/compound and therefore analyse their structure similarly to the structure of simple adpositions. Consider two adpositional phrases in (64).

(64) a. علی ساری رات زیر آسمان سویا.

ɑli	sari	rat	zer	asman	so-ya
Ali.M.3Sg	all	night.F.3Sg	under	sky.M	sleep-Perf.M.3Sg

‘Ali slept all the night under the sky.’

b. ندا نے پروفیسر کے زیر نظر ایک مقالہ لکھا.

nɪda=ne	pɑrofɛsɑr=ke	zer	nɑzɑr	ek
Nida.F.3Sg=Erg	professor.3Sg=Gen.M.3Sg.Obl	under	supervision.F	one

mɑqɑlɑh	lɪk^{h}-a
article.M.3Sg	write-Perf.M.3Sg

‘Nida wrote an article under the supervision of a professor.’

The f-structure representation of a simple adpositional phrase *zer asman* ‘under the sky’ in (64a) is given in Figure 6.12. And if the complex adposition *zer nazar* ‘under supervision (of)’ is not considered as complex then the f-structure of the adpositional phrase *parofɛsar=ke zer nazar* ‘under supervision of a professor’ used in (64b) should look as in Figure 6.13.

$$\begin{bmatrix} \text{PRED} & \text{‘zer}\left\langle \text{asman} \right\rangle\text{’} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{‘asman’} \\ \text{PERS} & \text{3, NUM sg, GEND masc, CASE nom} \end{bmatrix} \end{bmatrix}$$

Figure 6.12: Simple adposition

$$\begin{bmatrix} \text{PRED} & \text{‘zer}\left\langle \text{nɑzɑr} \right\rangle\text{’} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{‘nɑzɑr}\left\langle \text{pɑrofɛsɑr} \right\rangle\text{’} \\ \text{OBL} & \begin{bmatrix} \text{PRED} & \text{‘pɑrofɛsɑr’} \\ \text{PERS} & \text{3, NUM sg, CASE gen} \end{bmatrix} \\ \text{PERS} & \text{3, NUM sg, GEND fem, CASE nom} \end{bmatrix} \end{bmatrix}$$

Figure 6.13: An analysis of *zer nazar* ‘under supervision (of)’

To analyze constructions like *zer nazar* ‘under supervision (of)’ as in Figure 6.13, one would have to ensure the following:

- Only some nouns like *nazar* ‘supervision’ should not take premodifiers or post-modifiers when they become the object of the adposition *zer* ‘under’. Other examples of such nouns/objects can be seen in Table 6.8.

- The genitive marked complements of only such objects as *nazar* ‘supervision’ can be placed before the adposition *zer* ‘under’. This rules out other similar but ungrammatical structures like *nɪda=ke zer-e pa* ‘*under the foot of Nida’.

The last point would lead to a flat c-structure of such adpositional phrases as is shown in Figure 6.14 below. This analysis would also imply that there are discontinuous NPs within Urdu PPs as we see that the noun phrase *parofɛsar ki nazar* ‘supervision of professor’ is discontinuous. The noun *nazar* ‘supervision’ is on the right side of the adposition and its complement *parofɛsar ki* is one the left side of the adposition.

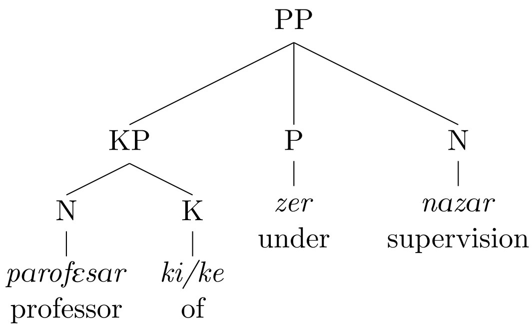
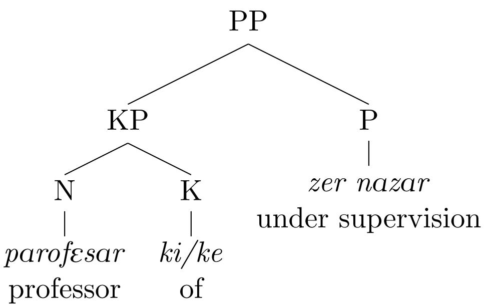


Figure 6.14: Flat structure for Urdu PPs

Another simple analysis could be that expressions like *zer nazar* ‘under supervision (of)’ are multiword lexical predicators which take genitive marked objects/complements in Urdu. The c-structure and the f-structure of the phrase *parofɛsar=ke zer nazar* ‘under supervision of a professor’ based on this analysis is given in Figure 6.15.



PRED	‘zer nazar⟨parofɛsar⟩’		
OBJ	PRED	‘parofɛsar’	
	PERS	3, NUM sg, CASE gen	

Figure 6.15: *zer nazar* as a multiword lexical adposition

As already mentioned, the syntactic differences of nouns in complex adpositions support an analysis in which the two lexical categories *zer* ‘under’ and *nazar* ‘supervision’ project to a complex adposition *zer nazar* ‘under supervision (of)’, which heads the entire phrase. The lexically selected complement of the relational noun in such phrases

is syntactically realized as the object of the complex adposition. So, the phenomeon of complex adposition formation is somehow similar to verbal complex formation although the former is not very productive. Now I move to the linguistically motivated implementation of complex adpositions in LFG.

Alsina (1996) and Butt (1995) have argued that complex predicate formation should take place in the syntax. Frank (1996), however, takes the opposite position and places complex predicate formation within the lexical component. Although in her approach subcategorization requirements and restrictions on combinations are specified within the lexical entries, still the separate pieces are combined in the syntax. Butt *et al.* (2003) implemented the syntactic combination of complex predicates of in Urdu via the RESTRICTION operator (Kaplan & Wedekind, 1993).

The above proposals in LFG have been posited for the verbal complex formation in different languages. Trawinski (2003) proposed an HPSG-based linguistically motivated, formal treatment of complex prepositions, applicable for computational platforms intended for developing typed feature structure grammars. Following Butt *et al.* (2003), I posit the proposal for licensing the complex adpositions in the LFG formalism (Raza, 2011) via the RESTRICTION operator. This operator is used to add and/or delete an argument of a predicator.

The subcategorization requirement and the features recognizing some specific combinations of adpositions and nouns are included in lexical entries and are checked at the syntactic level. The syntactic composition of two predicates to yield one complex predicator is done using RESTRICTION. The complement of the noun part is suppressed and is assigned as the object of the complex adposition. The lexical entry for the basic adposition *zer* 'under' and the grammar rule constructing the complex adposition is given in Figure 6.16.

I posit two entries for the adposition *zer* 'under'. The first entry states that it takes a variable argument which can be filled in the syntax. The %Arg is an XLE notation used for a variable. This entry permits the formation of complex predicators. The second entry is the basic entry where the adposition is just used as a simple adposition. The grammar rule states that the PRED of the noun is assigned to the variable argument of the adposition and the argument of the noun becomes OBJ of the newly constructed adposition.

The c-structure and the f-structure of the syntactically constructed complex adposition *zer nazar* 'under supervision (of)' is shown in Figures 6.17 and 6.18 respectively. In the f-structure, we see that the complement *parofɛsar* 'professor' of the noun *nazar* 'supervision' has become the object of the complex adposition.

Lexical Entry

zer P ∗ {(↑ PRED) = 'zer <%ARG1>';
|(↑ PRED) = 'zer <↑ OBJ>';}

Grammar Rule

Pc → {P: ↑ = ↓;
N: ↓\PRED\OBL = ↑\PRED\OBJ
(↑ PRED ARG1) = (↓ PRED)
(↑ OBJ) = (↓ OBL)
| ...
| ...
| ...}

Figure 6.16: An excerpt from the grammar rules

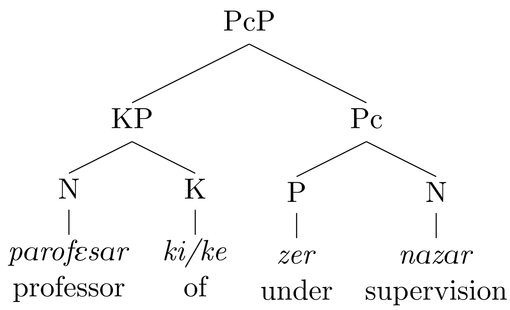


Figure 6.17: The c-structure of syntactically constructed complex adposition

$$\begin{bmatrix} \text{PRED} & \text{`zer}\left\langle \text{nazar}\left\langle \text{parofɛsar}\right\rangle\right\rangle\text{'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`parofɛsar'} \\ \text{PERS} & \text{3, NUM sg, CASE gen} \end{bmatrix} \end{bmatrix}$$

Figure 6.18: The f-structure of syntactically constructed complex adposition

6.6 Complex adpositions to simple adpositions

Complex adpositions comprising of more than one morpheme in a grammaticalization process may later be reduced into single word. Condensing is one form of reduction in which two morphemes are combined into one (Kortmann & König, 1992), for example, *beside* from *be sidan* and *instead* from *in steed*. Examples of condensing in Urdu/Persian are *bah qɔl* to *baqɔl* 'per saying' and *bah tɔr* 'by way' to *batɔr*.

Another form of reduction is noted in Urdu and Persian (Raza, 2010d), in which the

basic preposition (Persian) or the case marker (Urdu) is dropped and the noun itself is used as an adposition. In Urdu, a noun with some case marker acts first like a complex adposition, and then with the passage of time the case marker is dropped/lost and the bare oblique form of the noun itself acts like a simple adposition.[4] Some examples of this phenomenon are given in Table 6.9.

Complex adposition	**Simple adposition**
(ke) bal par 'with support/power (of)'	*(ke) bal*
(ke) vaqt par 'at the time (of)'	*(ke) vaqt*
(ke) ba'ıs se 'because (of)'	*(ke) ba'ıs*
(ke) bahane se 'by excuse (of)'	*(ke) bahane*
(ke) raste se 'through way (of)'	*(ke) raste*
(ke) zari'e se 'by means (of)'	*(ke) zari'e*
(ke) sabab se 'because (of)'	*(ke) sabab*
(ke) tʊfɛl se 'by grace (of)'	*(ke) tʊfɛl*
(ki) tarah se 'similar (of)'	*(ki) tarah*
(ke) bare mẽ 'about'	*(ke) bare*
(ke) natije mẽ 'in result of'	*(ke) natije*
(ke) dɔran mẽ 'during'	*(ke) dɔran*
(ki) surat mẽ 'in case (of)'	*(ki) surat*

Table 6.9: Complex adpositions to simple adpositions

In contemporary Urdu *bare* 'about' and *dɔran* 'during' are only used as adpositions, although originally in Persian, these words are nouns used in the meaning of 'regard' and 'period'. Another noun *dɔranıyah* is used in Urdu for the meaning of 'period'. However most of other lexical items except the case markers in Table 6.9 are used as nouns as well.[5]

6.7 Still more adpositions

In addition to all of the above adpositions, there are some further adpositions/adpositional phrases which are described in this section.

[4]In some adverbial constructions, the case marker can also optionally be dropped as in *gʰanṭõ (tak)* 'for hours'.

[5]It is worth mentioning here that sometimes, in case of infinitival clause complements, the adposition *lıye* 'for' together with the genitive case marker *ke* is suppressed and only the infinitival clause (the infinitive being in masculine oblique form) encodes the adpositional meaning as is shown in (65), where *kıtab xarid-ne* means 'for purchasing the book'.

(65) ندا کتاب خریدنے (کے لیے) گئی ہے.

nıda	kıtab	xarid-ne(=ke	lıye)	ga-i	hɛ
Nida.F	book.F	buy-Inf.Obl(=Gen	for)	go-Perf.F.3Sg	be.Pres.3Sg

'Nida has gone to purchase the book.'

6.7.1 *hat^{h}õ* 'hands'

The adposition *hat^{h}õ* is plausibly the reduced form of *hat^{h}õ se/mẽ*. It is used in three senses as in (66) below.

(66) a. سوات کا شہر دہشت گردوں کے ہاتھوں تباہ ہوا.

sɑvat=ka	ʃɑhr	dɑhʃɑt-gɑrdõ=ke	hat^{h}õ	tɑbah
Sawat=Gen.M.Sg	city.M.Sg	terrorist.3Pl=Gen.3Pl	hand.3Pl.Obl	destroyed

hu-a
be-Perf.M.Sg
'The city of Sawat was destroyed by terrorists.'

b. علی نے احمد کے ہاتھوں اپنا مکان بیچا.

ɑli=ne	ɑhmɑd=ke	hat^{h}õ	ɑpna	mɑkan	bec-a
Ali.M=Erg	Ahmad.M=Gen.3Pl	hand.3Pl	GenR.M.3Sg	house.M	sell-Perf.M.Sg

'Ali sold his house to Ahmad.'

c. علی نے احمد کے ہاتھوں ندا کو خط بھیجا.

ɑli=ne	ɑhmɑd=ke	hat^{h}õ	nɪda=ko	xɑt	b^{h}ej-a
Ali.M=Erg	Ahmad.M=Gen.3Pl	hand.3Pl	Nida.F=Dat	letter.M	send-Perf.M.Sg

'Ali sent a letter to Nida through Ahmad.'

In (66a), the complement of the adposition encodes the agent of the sentential clause, in (66b) its complement encodes a goal argument and in (66c) the complement of the adposition encodes an intermediary/supporting agent.

6.7.2 *le kar* 'from'

The canonical use of the subordinating clause conjunction *kar* is illustrated in (67) below. This conjunction is used when two actions are done in a sequence. The word combination *le kar* is canonically used in (67b), however, it is is not used in the canonical sense in (68).

(67) a. ندا کھانا کھا کر سو گئی.
nɪda [kʰana kʰa kɑr] so gɑ-i
Nida.F meal.M eat CConj sleep go-Perf.F.3Sg
'Nida went to sleep after having eaten meal.'

b. علی اپنی کتاب ندا سے لے کر چلا گیا.
ɑli [ɑpni kɪtab nɪda=se le kɑr] cɑl-a gɑ-ya
Ali.M GRefl.F book.F Nida.F=Abl take Conj.having walk-Perf.M go-Perf.M
'Ali went after having taken his book from Nida.'

(68) a. ندا نے صبح سے لے کر شام تک کتاب پڑھی.
nɪda=ne sʊbʊh=se le kɑr ʃam=tak
Nida.F=Erg morning.F=Abl take Conj.having evening.F=Temp.till
kɪtab pɑɽʰ-i
book.F.Sg read-Perf.F.Sg
'Nida read a book from morning to evening.'

b. علی نے حاصلوالا سے لے کر ملتان تک سفر کیا.
ɑli=ne hasɪlvala=se le kɑr mʊltan=tak sɑfɑr
Ali.M=Erg Hasilwala=Abl take Conj.having Multan=Loc.till journey.M
kɪ-ya
make-Perf.M
'Ali travelled from Hasilwala to Multan.'

The expression *le kar* as an adposition is used to encode source time (68a) or source location (68b) in context of 'from-to' time or 'from-to' location.[6]

[6]The lexical sign *baje* is used to encode the time in full hour (69a) and when the time is mentioned in hours and minutes then hours are encoded with *baj kar* (69b).

(69) a. ندا ۸ بجے سو گئی.
nɪda 8 bɑje so gɑ-i
Nida.F.3Sg 8 O' clock sleep go-Perf.F.Sg
'Nida went to sleep at 8 O' clock.'

b. علی ۸ بج کر ۱۰ منٹ پر چلا گیا.
ɑli 8 bɑj kɑr 10 mɪnt=par cɑl-a go-Perf.M.3Sg
Ali.M 8 ring Conj.having 10 minute=Temp.on walk-Perf.M go-Perf.M.Sg
'Ali went at 10 past 8.'

6.7.3 Arabic prepositions in Urdu

A very few prepositions of Arabic are also used in Urdu, but they are not used productively. Some typical prepositional phrases of Arabic used in Urdu are given in Table 6.10.

Preposition	Examples of prepositional phrases
bɪ 'with'	*bɪ-l-ɪttɪfaq* 'unanimously', *bɪ-l-tartib* 'in order'
bɛn 'between'	*bɛn-ʊl-adyan* 'between religions', *bɛn-al-satur* 'between lines'
ɑla 'on'	*ɑla-l-ɪ'lan* 'openly', *ɑla-l-amum* 'generally'
fi 'in'	*fɪ-l-asl* 'in fact', *fɪ-l-hal* 'presently'
ma' 'with'	*ma' tarjamah* 'with translation', *ma' tafrih* 'with explanation'
mɪn 'from'	*mɪn ba'd* 'after', *mɪn jʊmlah* 'in all'
mɪn janɪb 'from'	*mɪn janɪb allah* 'from God' , *mɪn janɪb zɛd* 'from Zaid'
mɪn hɛs 'as'	*mɪn hɛs-ʊl-qɔm* 'as a nation', *mɪn hɛs-ʊl-jama'at* 'as a party'

Table 6.10: Arabic prepositional phrases in Urdu

In contrast to Persian loaned prepositions, which are used productively in Urdu, Arabic prepositions are not used productively in Urdu.

6.8 Classes of adpositions

Based on marking of case on the complements of adpositions, Urdu adpositions can be divided into four classes.

Class 1: There is one class of adposiitons in Urdu which take only genitive marked complements. Some examples are listed below:

- *nɪda=ke saṭh* 'accompanied by Nida'
- *nɪda=ki babat* 'about Nida'

Class 2: In another class, adpositions take either nominative or genitive marked complements. An example is given below.

- *nɪda baɣɛr* 'without Nida'
- *nɪda=ke baɣɛr* 'without Nida'

Class 3: In the third class, adpositions take either genitive or ablative marked arguments. An example is given in the following.

- *nıda=ke age* 'before Nida' (in context of position in space)
- *nıda=se age* 'ahead of Nida'

Class 4: In this class, adpositions take only ablative marked arguments. For example:

- *nıda=se pahle* 'before Nida' (in context of time)

Some Persian loaned adpositions can take an *ezafe* linked noun or a prepositional phrase as their complements. If complements in the *ezafe* construction and preposional phrases as complements are considered in the classification then the above classes may further be subdivided into subclasses.

6.9 Conclusion

In this chapter different predicators used as adpositions in Urdu were explored. Th compound adposition analysis of spatail adpositions with the genitive marker was opposed. It was shown that spatial adpositions in Urdu can be explained by vector space semantics and that the hierarchy of spatial adpostions in Urdu is the mirror image of the hierarchy of English prepositions as devised by Svenonius with the exception of the position of the Deg and Deix heads. A model of spatial adpositions was proposed in LFG in terms of LEX-SEM features drawing on Svenonius' notions of spatial expressions which themselves are influenced by vector space semantics.

Argumentation was made why some predicators in Urdu are complex adpositions. The noun part in complex adpositions show different syntactic distribution from its syntactic distribution elsewhere. With some complex adpositions formed with the basic adposition *zer* 'under', the hierarchical c-structure of adpostional phrases is not possible because the genitive complement of the noun part becomes non-contigious to the head noun. I therefore analysed them as complex adpostions and proposed to compose them at the level of syntax. Complex adpositions in LFG were implemented via RESTRICTION operator.

This detailed investigation of adpositonal predicators in Urdu can be useful in developing computing applications for Urdu language.

Chapter 7

Conclusions

In this thesis, I have focused on identifying and exploring different types of predicators and patterns of their subcategorization frames in Urdu. Not having at hand refined resources for Urdu, for example, part-of-speech tagged corpora or tree bank, it was investigated how to acquire subcategorization of verbs from a raw Urdu corpus. Challenges to automatic subcategorization acquisition from a raw Urdu corpus were discussed in Chapter 2. Urdu is a free word order language in which major constituents can scramble among each other in a clause. That is, an argument of a verb is not strictly bounded to some specific position in a sentential clause, although the verb itself usually comes last. So, identification of the argument based on its position is not possible in Urdu in contrast with other languages like English where some arguments can be identified only by their positions in a sentence.

The participants of verbs in Urdu sentences are usually marked for case by different case clitics. It first seemed straightforward to identify different arguments based on case marking. But it turned out that case clitics pose challenges. There is not a one-to-one correspondence between the case clitic form and the case feature. That is, the same clitic form is used to mark nouns for more than one case. Another challenge is that the same grammatical function is marked for different cases in different situations.

Furthermore, both nouns and verbs can subcategorize for their arguments. So, it was also a challenge to identify which case phrase is semantically combined with which predicator. The case phrase attachment ambiguities in some sense resemble to the PP-attachment ambiguities in English and other langauges. To identify the complementizer clause based on the complementizer form is not trivial in Urdu as the same form of complementizer is also used for many other functions.

Having explored the above challenges, an algorithm was devised to acquire subcategorization information from a raw Urdu corpus. In the subcategorization acquisition system for Urdu (SASU) many of above challenges were addressed and a few were

ignored. The SASU system was presented in Chapter 3. The strategy of the SASU system differs from existing strategies in two respects. For one, lexical clues of case are used. Secondly, the frames are identified indirectly from the extracted case phrases by applying some meta rules. This system comprises of two input repositories and a verb conjugator and four more components. About 700 basic verbs of Urdu were collected from different resources and a corpus of Urdu obtained mainly from news sites was cleaned and segmented. The different morphological inflectional paradigms of verbs were analyzed and an Urdu verb-conjugator was implemented as a supplementary part of the SASU system.

To extract subcategorization frames of a verb, a component of the system finds all candidate sentences by comparing the conjugation forms of the verb with tokens of sentences in the corpora. Necessary screening of candidate sentences is made to make it sure that the verb in those sentences is used as a main verb and that it is not of the verb of some subordinating or coordinating clause. Another component then builds different case clitics and complementizer combinations and collect their frequencies in the corpus. The third component of the system filters out potentially invalid combinations based on statistical method. The final component induces the frames of verbs from the valid case clitics and complementizer combinations. Results of subcategorization frames of 60 basic verbs in a summarized form were reported. These 60 verbs were chosen based on sufficient number of their occurrences in the corpora.

Due to the diversified syntacto-semantic behavior of the basic verb *ho* 'be/become', it was not viable to extract its subcategorization information by the developed system. So, this verb was individually investigated in Chapter 4 for its different uses and subcategorization frames. It was shown that the verb *ho* can basically be classified into stative *ho* and dynamic *ho*. The syntactic distribution of stative *ho* and dynamic *ho* are different in terms of aspect, taking the light verb *ja* 'go' and making modal construction with the verb *cah* 'want'. Both stative *ho* and dynamic *ho* act either intransitively or a as a copula. Syntactic frames of both types of copula were explored. Dynamic copula only subcategories for a subset of the frames selected by the stative copula. Although Urdu is a free word order language, the position of the participant does matter for interpreting it as a subject or a complement in case of copular sentences. The characterizing participles which are constructed by perfect form of *ho* were also analyzed with respect to the arguments they modify.

Arguments of deverbal adjectives and deverbal nouns were investigated in Chapter 5. Noun phrases containing multiple instances of genitive elements were explored and the order of different genitive modifiers in them was established. It was shown that only attributive genitives can stack together at same level before adjectives in Urdu NPs,

otherwise there is always a hierarchical structure. Attributive genitives show syntactic distribution similar to adjectives. Furthermore it was shown that some nouns in Urdu can take two genitive marked arguments. A classification of nouns was made based on number and type of genitive marked argumetns. It was reported that discontinuous constituents are generated in Urdu NPs when an argument taking noun is modified by some argument taking adjective or if the argument of the head noun itself licenses its argument. Heads cannot appear before their argument in noun phrases. Argument-less adjectives are always contiguous to the head noun. The syntactic explanation of the phenomenon was provided in terms of multiple movements across different projections. In LFG a flat c-structure was proposed for Urdu NPs. A correct f-structure was generated by making use of different operators of XLE in the grammar rules.

Adpositions as predicators were analyzed in Chapter 6. A model of spatial adpositions in LFG was proposed in terms of lex-sem features drawing on Svenoniuous notions of spatial expressions. Different classes of adpositions were made based on the case of their complements. An evidence for complex adpositions in Urdu was provided. It was shown that nouns in complex adpositions show different syntactic distribution compared with their normal syntactic distribution. Linguistically motivated implementation of complex adpositions was presented.

To conclude, this thesis reports results of an exhaustive research made on patterns of predicators and their subcategorization frames in Urdu. An acquisition system is presented which extracts the subcategorization frames of Urdu verbs based on lexical cues of case clitics and complementizer forms. Having a very large and balanced corpus of Urdu, the SASU system presented in the thesis can be used to build the broad-coverage lexicon of Urdu verbs enriched with their subcategorization information in terms of grammatical functions coupled with their case marking. This system can support identification of complex predicates and also useful in discovering different patterns of syntactic alternations for Urdu verbs. By adding more bits of adpositions in the information vector, the adpositional arguments of verbs can also be extracted.

As a future work, the system can be generalized for South-Asian languages as many South-Asian languages like Saraiki and Sindhi are structurally very close to Urdu. The language selection and other parameters could be set at the interface level and subcategorization information in different South-Asian languages can be acquired by using the same inside technology. More fine-grained and advanced computing applications for South-Asian languages in general and particularly Urdu can be developed by using the SASU system as a core module.

The classes of verbs based on their syntactic frames can be explored. Incorporating the information of syntactic frames with allowed alternations and features of selectional

preferences could be more useful in exploring semantic classes of verbs in Urdu. Formal analysis needs to be worked out for phenomena like even predicates and many complex predicates reported in the thesis which have not yet been analyzed and implemented in context of some formal theory.

Zusammenfasung

In dieser Dissertation habe ich mich mit der Identifikation und Untersuchung verschieden Typen der Prädikation und deren Muster der Subkategorisierung beschäftigt. Da keine präzisen Resourcen wie zum Beispiel part-of-speech getaggte Korpora oder eine Baumbank für Urdu verfügbar sind, wurde untersucht, wie man die Subkategorisierungsrahmen von Verben aus rohen Urdu Korpora erfassen kann. Die Herausforderungen der automatischen Erfassung von Subkategorisierungsrahmen aus einem rohen Korpus für Urdu werden im zweiten Kapitel diskutiert. Urdu ist eine Sprache mit freier Wortstellung, bei der sich die Hauptkonstituenten frei im Satz bewegen können. Das bedeutet, dass ein Verbargument nicht strikt an eine Position im Satz gebunden ist, obwohl das Verb selber normalerweise am Schluss steht. Demnach ist die Identifikation eines Arguments aufgrund seiner Position in Satz im Falle von Urdu nicht möglich, dies steht im Gegensatz zu anderen Sprachen wie Englisch, wo einige Argumente aufgrund ihrer Position im Satz bestimmt werden können.

Argumente von Verben werden in Urdu Sätzen generell mit Klitika kasusmarkiert. Es sah zunächst relativ unkompliziert aus, diese Argument aufgrund ihrer Kasusmarkierung zu identifizieren. Allerdings stellte sich heraus, dass diese Kasusklitika einige Herausforderungen aufwerfen. Es gibt keine Eins-zu-eins-Übereinstimmung zwischen Kasusklitikon und Kasuseigenschaft. Das bedeutet, dass dasselbe Klitikon Nomen mit mehr als einem Kasus markieren kann. Eine weitere Herausforderung ist, dass dieselbe grammatische Funktion, abhängig von der jeweiligen Situation, mit unterschiedlichem Kasus versehen sein kann.

Darüber hinaus können Nomen und Verben Argumente haben. Deswegen war es auch eine Herausforderung herauszufinden, welche kasusmarkierte Phrase semantisch mit welchem Prädikat verbunden ist. Die Bindungsambiguitäten der Kasusphrase sind in etwa vergleichbar mit der Bindungsambiguität von Präpositionalphrasen im Englischen und anderen Sprachen. Die

Identifikation der Komplementiererphrase auf der Basis des Komplementierers ist nicht einfach in Urdu, da dieselbe Komplementiererform auch für andere Funktionen verwendet wird.

Nach der Erforschung der oben genannten Herausforderungen wurde ein Algorithmus konzipiert, der die Subkategorisierungsinformation aus dem rohen Urdu Korpus gewinnt. In diesem System zur Gewinnung von Subkategorisierungsrahmen (SASU) sind viele der oben genannten Herausforderungen berücksichtigt, einige wurden auch ignoriert. Das SASU System wurde in Kapitel 3 präsentiert. Die Herangehensweise des SASU Systems unterscheidet sich von bereits existieren Strategien in zweierlei Hinsicht. Zum einen werden lexikalische Hinweise für Kasus benutzt. Zweitens werden die Rahmen indirekt aus den extrahierten Kasusphrasen identifiziert, indem Metaregeln angewandt werden. Dieses System beinhaltet zwei Archive für die Eingabe, ein Verbkonjugator und vier weitere Komponenten. Rund 700 Basisverben in Urdu wurden von unterschiedlichen Quellen gesammelt, zudem wurde ein Urdu Korpus auf der Basis von verschiedenen Nachrichtenseiten aufgeräumt und segmentiert. Die verschiedenen morphologischen Flektionsparadigmen der Verben wurden analysiert und ein Verbkonjugator wurde als unterstützender Teil des SASU Systems implementiert.

Um die Subkategorisierungsrahmen eines Verbes zu extrahieren findet eine Komponente des Systems alle Kandidatensätze in dem die Konjugationsformen des Verbes mit den Token der Sätze im Korpus überstimmen. Um sicherzugehen, dass das Verb in diesem Sätzen als Hauptverb benutzt wird und nicht als Verb eines Nebensatzes oder einer Koordinierungskonstruktion, ist ein Durchsehen der Kandidatensätze nötig. Eine weitere Komponente baut dann verschiedene Kasusklitika und Komplementiererkonstruktionen zusammen und ermittelt deren Häufigkeit im Korpus. Die dritte Komponente des Systems filtert auf der Basis einer statistischen Komponente potentiell falsche Kombinationen aus. Die letzte Komponente induziert die Subkategorisierungsrahmen der Verben aufgrund der validen Kombinationen von Kasusklitika und Komplementiererkombinationen. Die Ergebnisse der Subkategorisierungsrahmen von sechzig Basisverben werden in zusammengefasster Form in der Dissertation dargestellt. Die Auswahl dieser Verben wurde aufgrund ihres ausreichenden Vorkommens in den Korpora getroffen.

Das Verb *ho* 'to be/become' wurde aufgrund seines breit gefächerten syntaktisch-semantischen Verhaltens separat in Kapitel 4 behandelt. Es wurde gezeigt, dass das Verb ho generell in statisches *ho* und dynamisches *ho* eingeteilt werden kann. Die syntaktische Verteilung des statischen und dynamischen *ho* unterscheiden sich mit Hinsicht auf ihren Aspekt, ihrer Kombination mit dem Hilfsverb *ja* 'gehen' und der Fähigkeit zu Modalkonstruktionen mit dem Verb *cah* 'wollen'. Statisches und dynamisches *ho* kommen beide entweder in einer Intransitivkonstruktion oder in einer Kopulakonstruktion vor. Die syntaktischen Rahmen für beide Arten der Kopulakonstruktion wurden untersucht. Das dynamische *ho* subkategorisiert nur für eine Untermenge an syntaktischen Rahmen die für das statische ho gelten. Obwohl Urdu eine Sprache mit freier Wortstellung ist, ist die Position des Arguments in Kopulakonstruktionen wichtig für seine Interpretation als Subjekt oder Komplement. Die charakterisierenden Partizipien, die mit der Perfektform von ho geformt werden, werden ebenfalls mit Hinsicht auf die von ihnen modifizierten Argumenten analysiert.

Argumente von Verbaladjektiven und Verbalnomen wurden in Kapitel 5 untersucht. Nominalphrasen mit mehreren Genitivelementen wurden analysiert und es wurde gezeigt, dass nur attributive Genitive auf der gleichen Ebene der Urdu NP zusammengesetzt werden können, in allen anderen Fällen gibt es eine hierarchische Struktur. Weiterführend wurde gezeigt, dass einige Nomen in Urdu zwei genitiv-markierte Argumente besitzen können. Aufbauend auf der Anzahl und des Typs der genitiv-markierten Argumente wurde eine Nomenklassifizierung durchgeführt. Dabei wurde berichtet, dass nicht-kontinuierliche Konstituenten in der Nominalphrase in Urdu dadurch gebildet werden, dass ein subkategorisierendes Nomen von einem subkategorisierenden Adjektiv modifiziert wird oder dass das Argument des Nomes ein eigenes Nomen lizensiert. The syntaktische Erklärung des Phänomens wird bereitgestellt, ebenso wie eine Implementierung im Rahmen von LFG.

Adpositionen als Prädikatoren wurden in Kapitel 6 analysiert. Ein Modell für räumliche Adpositionen im Rahmen von LFG wurde vorgeschlagen, auf der Basis der von Svnenonius lexikalisch-semantischen Ideen zum Ausdruck von Räumlichkeit. Auf der Basis des Kasus ihrer Objekte wurden Adpositionen in verschiedene Klassen eingeteilt. Es wurde gezeigt, dass Nomen in komplexen Adpositionen eine unterschiedliche syntaktische Verteilung

zur ihrer normalen syntaktischen Verteilung aufweisen. Die linguistisch motivierte Implementation dieser komplexen Adpositionen wurde gezeigt.

Zusammenfassend berichtet diese Dissertation über die Ergebnisse umfassender Forschung zu den Mustern von Predikatoren und deren Subkategorisierungsrahmen in Urdu. Durch einen sehr groβen, ausreichend gewichteten Korpus kann das SASU System, das in dieser Dissertation vorgestellt wurde, benutzt werden, um ein umfassendes Verblexikon mit Subkategorisierungsinformation im Sinne ihrer grammatischen Funktionen und deren Kasusmarkierung zu erstellen. Dieses System kann für die Identifikation von komplexen Prädikaten verwendet werden, sowie bei der Erkennung von syntaktischen Alternationen von Urdu Verben behilflich sein. Als Teil der zukünftigen Aufgaben können Verbklassen aufgrund der syntaktischen Rahmen untersucht werden. Die Inkorporation der syntaktischen Rahmen mit möglichen Alternationen und Eigenschaften der lexikalischen Beschränkungen können bei der Untersuchung von semantischen Verbklassen in Urdu eingesetzt werden. Die formale Analyse kann insoweit ausgebaut werden, als dass für manche Phänomene von Verben und komplexen Prädikaten, über die in dieser Dissertation berichtet wurde, formal weder eine Analyse noch eine Implementation existiert.

References

Agbayani, B. & Golston, C. (2005). Movements of nonconstituents and insensitivity to islands in Ancient Greek. Rutgers Optimality Archive. 183

Ahmed, T. (2006). Spatial, temporal and structural uses of urdu ko. In M. Butt & T. King, eds., *LFG06*. 38

Ahmed, T. (2011). *Spatial Expressions and Case in South Asian Languages*. Ph.D. thesis, University of Konstanz. 39, 42

Al-Sharifi, B. & Sadler, L. (2009). The adjectival construct in Arabic. In M. Butt & T.H. King, eds., *Proceedings of the LFG09 conference*, 26–43, Stanford: CSLI Publications. 178

Alsina, A. (1996). *The Role of Argument Structure in Grammar*. Stanford: CSLI Publications. 252

Anderson, M. (1983). Prenominal Genitive NPs. *Linguistics Review*, **3**, 1–24. 162

Anderson, S.R. (1977). On mechanisms by which languages become ergative. In C. Li, ed., *Mechanisms of Language Change*, 317–363, University of Texas Press, Austin, Texas. 37

Androutsopoulou, A. (2000). On the (In)ability of Prenominal Attributive Adjectives to Take Complements. In Billerey & Lillehaugen, eds., *WCCFL Proceedings*, vol. 19, 29–42, MA: Cascadilaa Press, Somerville. 193

Attia, M. (2008). A Unified Analysis of Copula Constructions in LFG. In M. Butt & T.H. King, eds., *Proceedings of the LFG08 Conference*, CSLI Publications. 3, 105

Austin, P. (2001). Lexical Functional Grammar. In *International Encyclopedia of the Social and Behavioral Sciences*, 8748–8754, Elsevier. 8

Austin, P. & Bresnan, J. (1996). Non-configurationality in Australian Aboriginal languages. *Natural Language and Linguistic Theory*, **14.2**, 15–68. 182, 183

Avgustinova, T. & Uszkoreit, H. (2003). Reconsidering the relations in constructions with non-verbal predicates. In *Proceedings of the 4th European conference on Formal Description of Slavic Languages (FDSL4)*. 111

Babby, L.H. (1997). Nominalization in Russian. In W. Browne, E. Dornisch, N. Kondrashova & D. Zec, eds., *Formal Approaches to Slavic Linguistics*, 54–83, Ann Arbor: Michigan Slavic Publications. 162

Baker, C. (1995). *Possessive Descriptions*. CSLI Publications, California. 158

Bartsch, R. (1986/87). The Construction of Properties under Perspectives. *Journal of Semantics*, 293–320. 178

Beames, J. (1879). *A comparative grammar of the modern Aryan languages of India*, vol. 3 vols. Trübner and Co., London, Reprinted by Munshiram Manoharlal, New Delhi in 1996. 148

Bessler, P. (1992). Some Properties of Attribution in English and German. In *Proceedings of the 1992 Annual Conference of the Canadian Linguistic Association*, 15–28. 193

Bhatt, R. (1998). Complex Predicates as Polarity Items. 41

Bhatt, R. (2003). *Causative derivations in Hindi*. 19

Bhatt, R.K. (2005). Long-distance agreement in Hindi-Urdu. *Natural Language and Linguistic Theory*, 757–807. 68, 162

Bögel, T. & Butt, M. (to appear). Possessive Clitics and Ezafe in Urdu. In K. Börjas & D. Denison, eds., *Morpho-syntactic categories and the expression of possession*, John Benjamins. 31

Boguraev, B. & Briscoe, T. (1989). Computatnal lexicography for natural language processing. Longman, London. 15

Bouchard, D. (1997). The Distribution and Interpretation of Adjectives in French: A consequence of Bare Phrase Structure. unpublished ms., Université du Quebéc á Montréal. 193

Bošković, Z. (2005). On the localitiy of left branch extraction and the structure of NP. *Studia Linguistica*, **59**, 1–45. 183

Brent, M. (1991). Automatic acquisition of subcategorization frames from untagged text. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, 209–214, Berkeley, CA. 14

Brent, M.R. (1993). From grammar to lexicon: Unsupervised learning of lexical syntax. *Computational Linguistics (19)2*, 243–262. 14, 56, 58, 61, 88, 89, 91

Bresnan, J. (1982). *The Mental Representation of Grammatical Relations*. MIT Press, Cambridge, Ma. 6, 10

Briscoe, T. & Carrol, J. (1993). Generalized probabilistic LR parsing for unification-based grammars. *Computational Linguistics (19)1*, 25–60. 4, 56

Briscoe, T. & Carrol, J. (1997). Automatic extraction of subcategorization from corpora. In *Proceedings of the 5th ACL Conference on Applied Natural Language Processing*, 356–363, Washington DC. 14, 56, 61, 89

Butt, M. (1993). Object Specificity and Agreement in Hindi/Urdu. In *Papers from the 29th Regional Meeting of the Chicago Linguistic Society*, 80–103. 28, 45

Butt, M. (1995). *The Structure of Complex Predicates in Urdu*. Stanford: CSLI Publications. 26, 68, 130, 162, 252

Butt, M. (1997). Aspectual Complex Predicates, Passives and Disposition/Ability. 41

Butt, M. (2005b). Restriction for Morphological Valency Alternations: The Urdu Causative. In M. Butt & T.H. King, eds., *Proceedings of the LFG05 Conference*. 19

Butt, M. (2006). The dative-ergative connection. In *Proceedings of the Colloque Syntax-Semantique Paris (CSSP) 2005*. 47

Butt, M. & King, T.H. (2005). The Status of Case. In V. Dayal & A. Mahajan, eds., *Clause Structure in South Asian Languages*, 153–198, Berlin: Springer Verlag. 19, 36, 37, 43, 44, 45, 114, 149, 189, 201, 212

Butt, M. & King, T.H. (2007). Urdu in a Parallel Grammar Development Environment. *Language Resources and Evaluation, Special issue on Asian Language Processing: State of the Art Resources and Processing*, 191–207. 170

Butt, M. & Rizvi, J. (2010). Tense and Aspect in Urdu. In P. Cabredo-Hofherr & B. Laca, eds., *Layers of Aspect*, 43–66, Stanford: CSLI Publications. 117

Butt, M., Fortmann, C. & Rohrer, C. (1996). Syntactic Analyses for Parallel Grammars: Auxiliaries and Genitive NPs. In *Proceedings of COLING 1996*, 182–187. 170

Butt, M., King, T.H., Nino, M.E. & Segond, F. (1999). *A Grammar Writer's Cookbook*. Stanford: CSLI Publications. 9, 10, 12, 26, 104, 182

Butt, M., Dyvik, H., King, T.H., Masuichi, H. & Rohrer, C. (2002). The parallel grammar project. In *Proceedings of COLING2002, Workshop on Grammar Engineering and Evaluation*, 1–7. 12

Butt, M., King, T.H. & Maxwell III, J.T. (2003). Complex Predicates via Restriction. In *Proceedings of the LFG03 Conference*, 92–104, Stanford: CSLI Publications. 252

Campbell, L. (1985). *The Pipl Language of El Salvador*. Berlin: Mouton. 102

Carlson, G. & Pelletier, F. (1995). *The Generic Book*. Chicago: University of Chicago Press. 121

Carlson, G.N. (1977). *Reference to Kinds in English*. Ph.D. thesis, University of Massachusetts at Amherst. 112

Carrol, J., Minnen, G. & Briscoe, T. (1998). Can Subcategorization Probabilities help a Statistical Parser? In *Proceedings of the 6th ACL/SIGDAT Workshop on Very Large Corpora*, 118–126, Montreal, Canada. 56, 61

Ćavar, D. (1999). *Aspects of the Syntax-Phonology Interface*. Ph.D. thesis, University of Potsdam. 183

Chakrabarti, D., Bhattacharyya, P. & Malshe, M. (2002). Examining the Syntactic Alternations of Hindi Verbs with Reference to Morphological Paradigm. In *Language Engineering Conference (LEC2002)*, Hyderabad, India. 23

Chime, Z. (2006). An account for Compound Prepositions in Farsi. In *Proceedings of COLING/AICL*, Sydney. 238, 241

Chomsky, N. (1970). Remarks on Nominalization. In R. Jacobs & P. Rosenbaum, eds., *Readings in English Transformational Grammar*, Ginn and Co., Waltham, Ma. 145

Comrie, B. (1997). The typology of predicate case marking. In J. Haiman, J. Bybee & S.A. Thompson, eds., *Essays on Language Function and Language Type: Dedicated to T. Givon*, Amsterdam: John Benjamins. 102

Curnow, T.J. (2000). Towards a cross-linguistic typology of copula constructions. In J. Henderson, ed., *Proceedings of the 1999 Conference of the Australian Linguistic society*. 102, 116

Dalrymple, M. (2001). Lexical Functional Grammar. *Syntax and Semantics. New York: Acasemic Press*, **34**. 8, 10

Davison, A. (1988). Constituent Structure and the Realization of Agreement Features. In G.L. L. Macleod & D. Brentari, eds., *Papers from the 24th Regional Meeting of the Chicago Linguistics Society*, 41–53, Chicago Linguistics Society, Chicago. 68

Davison, A. (1990). Peculiar passives. *Language 56(1)*, 42–66. 41

Davison, A.L. (2008). On the categorical identity of infinitives in Hindi/Urdu. In *South Asian Language Analysis Roundtable, South Asia Conference*. 162

Dayal, V. (1992). The Singular-Plural Distinction in Hindi Generics. In C. Barker & D. Dowty, eds., *Proceedings of the Second Conference on Semantics and Linguistics Theory, OSU Working Papers in Linguistics 40*, 39–58, Dept. of Linguistics, Ohio State University, Columbus, Ohio. 122

Dayal, V. (1994). Scope-marking as Indirect Wh-Dependency. *Natural Language Semantics*, 137–170. 187

Declerck, R. (1988). *Studies on Copula Sentences, Clefts and Pseudo-clefts*. Leuven: Leuven University Press and Foris. 102, 103

den Dikken, M. (2003). On the syntax of locative and directional adpositional phrases. Ms. CUNY. http.web.gc.cuny.edu./dept/lingu/dendikken/syntax_of_pp.pdf. 207

Déprez, V. (2003). *On the typology of syntactic positions and the nature of chains*. Ph.D. thesis, MIT, Cambridge, Ma. 184

Diesing, M. (1988). Bare Plural Subjects and the Stage/Individual Contract. In *Genericity in Natural Language Processings of the 1988 Tübingen Conference*. 112

Dipper, S. (2003). *Implementing and Documenting Large-Scale Grammars — German LFG*. Ph.D. thesis, IMS, University of Stuttgart. 182

DIXON, R. (1994). *Ergativity*. Cambridge: Cambridge University Press. 37

DOHERTY, C. (1996). Clausal structure and the Modern Irish copula. *Natural Language and Linguistic Theory*, 1–46. 102

DOWTY, D. (1991). THEMATIC PROTO-ROLES AND ARGUMENT SELECTION. *Language*, **67**, 547–619. 5

DOWTY, D.R. (1987). *Thematic Proto-Roles, Subject Selection, and Lexical Semantic Defaults*. Ohio State University, Columbus. 162

DOWTY, R.D. (1979). *Word meaning and Montague grammar*. Dordrecht: Kluwer Academic Publishers. 127

DUNNING, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics (19)1*, 61–47. 88

DWIVEDI, V. (1994). *Syntactic Dependencies and Relative Phrases in Hindi*. Ph.D. thesis, University of Massachusetts, Amherst. 187

EAGLES (1996). Recommendations on Subcategorization. 9, 14

EMBICK, D. (2004). On the structure of resultative participles in English. *Linguistic Inquiry*, **25**, 355–392. 139

EMONDS, J. (1976). *A Transformational Approach to English Syntax*. Academic Press, New York. 191

EVANS, R. & KILGARRRIFF, A. (1995). Mrds, standards, and how to do lexical engineering. In *Proceedings of the 2nd Language Engnineering Convention*, 125–134, London, England. 5

FANSELOW, G. & ĆAVAR, D. (2002). Distributed Deletion. In A. Alexiadou, ed., *Theoretical Approaches to Universals*, 65–107, Amsterdam, Benjamins. 183

FANSELOW, G. & FÉRY, C. (2006). Prosodic and Morphosyntactic Aspects of Discontinuous Noun Phrases: a Comparative Perspective. Ms. Potsdam University. 183, 185

FÉRY, C., PASLAWSKA, A. & FANSELOW, G. (2007). Discontinuous Constructions in Ukrainian. *Journal of Slavic Linguistics*, **15.1**. 183

FILLMORE, C.J. (1968). The case for case. In E. Bach & R.T. Harms, eds., *Universals in Linguistic Theory*, 1–88, Holt, Rinehart and Winston, New York. 5

Frajzyngier, Z. (1978). An analysis of be-passives. *Lingua*, **46(2)**, 133–156. 139

Frank, A. (1996). A note on complex predicate formation: Evidence from auxiliary selection, reflexivization, and past participle agreement in French and Italian. In M. Butt & T.H. King, eds., *The Proceedings of the LFG96 Conference*, Stanford: CSLI Publications. 252

Freeze, R. (1992). Existentials and other locatives. *Language*, **68.3**, 553–596. 105

Gahl, S. (1998). Automatic extraction of subcorpora based on subcategorization frames from a part-of-speech tagged corpus. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, 428–432, Montreal, Canada. 5

Gehrke, B. (2007). On directional readings of locative prepositions. In S. Blaho, L. Vicente & E. Schoorlemmer, eds., *Console XIV*, 99–120, Leiden: Student Organisation of Linguistics in Europe. 207

GENELEX (1993). Project Eureka GENELEX. Rapport sur la couche syntaxique. Technical Report Version 4.0, ASSTRIL, GSI-ERLI, IBM France and Sema Group, Paris. 15

Ghomeshi, J. (1997). Non-projecting nouns and the ezafe construction in Persian. *Natural Language and Linguistic Theory*, **15(4)**, 729–788. 31, 166

Giorgi, A. & Longobardi, G. (1991). *The Syntax of Noun Phrases*. Cambridge University Press, Cambridge. 192

Gouskova, M. (2001). Split Scrambling: Barriers as Violable Constraints. In K. Megerdoomian & L.A. Bar-el, eds., *WCCFL Proceedings*, 220–233, Somerville, MA: Cascadilla Press. 183

Grimshaw, J. (1990). *Argument Structure*. MIT Press, Cambridge, MA. 61, 162, 169

Grishman, R., Macleod, C. & Meyers, A. (1994). Comex syntax: building a computational lexicon. In *Proceedings of the International Conference on Computational Linguistics, COLING-94*, 268–272, Kyoto, Japan. 86

Gruber, J. (1976). *Studies in Lexical Realtions*. Ph.D. thesis, MIT. 5

Haider, H. (1985). Über sein oder nicht sein: zur Grammatik des Pronomens sich. In *Erklärende Syntax des Deutschen. [Studien zur deutschen Grammatick 25]*, 223–254, Tübingen: Narr. 184

Hale, K. (1980). *On the position of Warlpiri in a Typology of the Base*. Bloomington: Indiana University Linguistics Club. 182

Hale, K. (1983). Warlpiri and the grammar of non-configurational languages. *Natural Language and Linguistic Theory*, **1**, 5–47. 182

Hale, K. (1992). Basic Word Order in Two "Free Word Order" Languages. In *Pragmatics of Word Order Flexibility*, 63–82, John Benjamins, Amsterdam. 182

Haspelmath, M. (1994). Passive participles across languages. In B. Fox & P. Hopper, eds., *Voice, Form and Function*, 151–177, Amsterdam: John Benjamins. 139, 141

Helmantel, M. (2002). Interactions in the Duch adpositional domain. LOT, Utrecht. http://www.lotpublications.nl/publish/articles/000249/bookpart.pdf. 207

Hengeveld, K. (1992). *Non-verbal Predication: Theory, Typology, Diachrony*. Berlin: Mouton de Gruyter. 102

Hole, D. (2009). German free datives and Knight Move Binding. Universität Stuttgart. 109, 110, 179

Hurford, J.R. & Heasley, B. (1983). *Semantics: a course book*. Cambridge: Cambridge University Press, 1st edn. 2, 3

Huybregts, R. & van Riemsdijk, H. (2002). Location and locality. In M. van Oostendorp & E. Anagnostopoulou, eds., *Progress in grammar: articles at the 20th anniversary of the Comparison of Grammatical Models Group in Tilburg*, 1–23, Meertens Instituut, Amsterdam. 207

Ijaz, M. & Hussain, S. (2007). Corpus Based Urdu Lexicon Development. In *Conference on language and technology*, Peshawar. 16

Ikeya, A. (1995). Predicate-argument structure of English adjectives. In B.K. T'sou & T.B. Lai, eds., *Proceedings of the 10th Pacific Asia Conference on Language, Information and Computation*, 149–156. 178

Jackendoff, R. (1996). The architecture of the linguistic-spatial interface. In P. Bloom, M.A. Peterson, L. Nadel & M.F. Garrett, eds., *Language and Space*, 1–30, Cambridge, MA: The MIT Press. 222

Junger, J. (1981). Copula Constructions in Modern Hebrew. In T. Hoekstra, H. van der Hulst & M. Moortagt, eds., *Perspectives on Functional Grammar*, 117–134, Dordrecht: Foris. 102

KACHRU, Y. (1976). On the Semantics of the Causative Construction in Hindi-Urdu. In M. Shibatani, ed., *Syntax and Semantics 6: The Grammar of Causative Constructions*, 353–369, New York: Academic Press. 19

KAPLAN, R. & WEDEKIND, J. (1993). Restriction and correspondence-based translation. In *Proceedings of the Sixth European Conference of the Association for Computational Linguistics*, 193–202. 252

KAPLAN, R.M. & BRESNAN, J. (1995). Lexical-Functional Grammar: A Formal System for Grammatical Representations. In M. Dalrymple, R.M. Kaplan, I.M.J. T. & A. Zaenen, eds., *Formal issues in Lexical-Finctional Grammar*, 29–130, Stanford: CSLI Publications. 8

KAZENIN, K. (2005). A syntactic analysis of Split Constructions in Russian. Ms., University of Potsdam. 183

KHAN, G. (2005). *Jame'ul Qavaid, page 159*. 49, 235

KINYON, A. & PROLO, C.A. (2002). Identifying Verb Arguments and their Syntactic Function in the Penn Treebank. In *Proceedings of the 3rd International Conference on Languages Resources and Evaluation*, 1982–1987, Las Palmas de Gran Canaria, Spain. 14, 56

KIPPER-SCHULER, K. (2005). *VerbNet, A broad-coverage, comprehensive verb lexicon*. Ph.D. thesis, University of Pennsylvania. 14

KISS, K.E. (2002). *The syntax of Hungarian*. Cambridge University Press. 215

KOOPMAN, H. (1993). The Internal and External Distribution of Pronominal DP's. Ms. UCLA. 193

KOOPMAN, H. (1997). Prepositions, postpositions, circumpositions and particles: the structure of Dutch PPs. Ms. UCLA. http://www.linguistics.ucla.edu/people/koopman/papers/pp.pdf. 207

KORHONEN, A. (2002). Subcategorization acquisition. Ph.D. thesis published as Technical Report UCAM-CL-TR-530, University of Cambridge. 5

KORHONEN, A., GORREL, G. & MCCARTHY, D. (2000). Statistical Filtering and Subcategorization Frame Acquisition. In *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Lanaguage Processing and Very Large Corpora*, 199–205, Hong Kong, China. 86, 88

KORTMANN, B. & KÖNIG, E. (1992). Categorial reanalysis: the case of deverbal prepositions. *Linguistics*, **30**, 671–697. 253

KRATZER, A. (1995). Stage-Level and Individual-Level Predicates. In G. Carlson & F. Pelletier, eds., *The Generic Book*, Chicago University Press. 112

KUHN, J. (1998). Resource sensitivity in the syntax-semantics interface and the German split NP construction. In *Proceedings of the ESSLLI X Workshop on CurrentTopics in Constraint Based Theories of Germanic Syntax*, Saarbrücken. 183

KURZON, D. & ADLER, S. (2008). *Adpositions: pragmatic, semantic and syntactic perspectives*. Amsterdam: John Benjamins. 238

LAMARCHE, J. (1991). Problems for N°-Movement to Num-P. *Probus*, **3(2)**, 215–236. 193

LAPATA, M. (1999). Acquiring lexical generalizations from corpora: A case study for diathesis alternations. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, 397–404, Maryland. 86

LAPATA, M. & KELLER, F. (1998). Corpus Frequency as a Predictor of Verb Bias. A poster presented at AMLAP-98, Freiburg. 5

LAPATA, M., KELLER, F. & SCHULTE IM WALDE, S. (2001). Verb frame frequency as a predictor of verb bias. *Journal of Psycholinguistic Research*. 5

LARSON, R. (1997). LIN 550: Special Topics in Linguistics. Fall '97 Class Lecture Notes, SUNY at Stony Brook. 193

LARSON, R. (1998). Events and Modifications in Nominals. In *Proceedings from Semantics And Linguistic Theory (SALT) VIII*, Ithaca NY: Cornell University. 193

LEBEAUX, D. (1986). The Interpretation of Derived Nominals. In A.M. Farley, P.T. Farley & K.E. McCullogh, eds., *CLS (Papers from the General Session at the Twenty-Second Regional Meeting of the Chicago Linguistics Society)*, 22, 231–247. 164

LEGATE, J. (2002). *Warlpiri: Theoretical Implications*. Ph.D. thesis, MIT, Cambridge, Ma. 183

LEVIN, B. (1993). *English Verb Classes and Alternation: A Preliminary Investigation*. The University of Chicago Press, Chicago. 5, 14, 63

MAHAJAN, A. (1990). *The A/A-Bar Distinction and Movement Theory*. Ph.D. thesis, Massachussetts Institute of Technology. 37

MAHAJAN, A.K. (1989). Agreement and Agreement Phrases. In I. Laka & A.K. Mahajan, eds., *Functional Heads and Clause Structure, No. 10 in MIT Working Papers in Linguistics*, 217–252, MITWPL, Cambridge, MA. 68

MALING, J. (1983). Transitive adjectives: a case of categorial reanalysis. In F. Henry & B. Richards, eds., *Linguistic Categories: Auxiliaries and Related Puzzles*, vol. 1, 253–289, Dordrecht: D Reidel. 177, 191

MANNING, C.D. (1993). Automatic Acquisition of a Large Subcategorization Dictionary From Corpora. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*. 14, 56, 58, 61

MARÁCZ, L. (1989). Asymmetries in Hungarian. [Diss.] Rijksuniveriteit Gorningen. 215

MARCUS, M.P., MARCINKIEWICZ, M.A. & BEATRICE (1993). Building a Large Annotated Corpus of English. *Computational Linguistics (19)2*, 313–330. 56

MASICA, C.P. (1991). *The Indo-Aryan languages*. Cambridge: Cambridge University Press. 105, 110

MAXWELL, J.T. & KAPLAN, R.M. (1993). The interface between phrasal and functional constraints. *Cmputational Linguistics*, **19**, 571–589. 12

MCCLOSKEY, J. (2000). Quantifier Float and Wh-Movement in an Irish English. *Linguistic Inquiry*, **31**, 57–84. 184

MCGREGOR, R.S. (1972). *Outline of Hindi Grammar: with exercises*. Oxford: Clarendon Press. 111

MEYERS, A., MACLEOD, C. & GRISHMAN, R. (1996). Standardization of the complement adjunct distinction. In *Proceedings of the 7th EURALEX International Congress*, 141–150, Goteborg Sweden. 61, 62, 70

MILLER, P. (1992). *Clitics and Constituents in Phrase Structure Grammar*. Garland, New York. 149

MITTENDORF, I. & SADLER, L. (2008). NP would like to meet GF. In M. Butt & T.H. King, eds., *Proceedings of the LFG08 conference*, 373–393, Stanford: CSLI Publications. 178

MOHANAN, T. (1994). *Argument Structure in Hindi*. Stanford: CSLI Publications. 23, 26, 37, 39, 43, 106, 108, 110, 111, 112, 113, 130

MORRIS-JONES, J. (1931). *Welsh Syntax*. Cardiff: The University of Wales Press Board. 178

MOYNE, J. (1971). Reflexive and Emphatic. *Language*, **47**, 141–163. 31

MÜLLER, S. (2004). Continuous or Discontinuous Constituents? A Comparison between Syntactic Analyses for Constituent Order and Their Processing Systems. *Research on Language and Computation*, 209–257. 183, 184

NEDJALKOV, V.P. & JAXONTOV, S.J. (1988). The typology of resultative constructions. In *Typology of Resultative Constructions*, 4–62, Amsterdam: John Benjamins. 139

NTHELITOES, D. (2004). *Syntax of Elliptical and Discontinuous Nominals*. Master's thesis, UCLA. 183

O'DONOVAN, R., BURKE, M., CAHILL, A., GENABITH, J.V. & WAY, A. (2005). Large-Scale Induction and Evaluation of Lexical Resources from the Penn-II and Penn-III Treebanks. *Computational Linguistics (31)3*, 329–365. 14, 56

PANDHARIPANDE, R. (1981). Interface of Lexicon and Grammar: some problems in Hindi grammar. *Studies in Linguistic sciences*, 77–100. 112

PANTCHEVA, M. (2006). Persian preposition classes. In P. Svenonius & M. Pantcheva, eds., *Nordlyd: Tromsø Working Papers on Language and Linguistics: 33.1, Special Issue on Adpositions*, 1–25, University of Tromsø, Tromsø. 223, 244

PANTCHEVA, M. (2008). The place of PLACE in Persian. In A. Asbury, J. Dotlacil, B. Gehrke & R. Nouwen, eds., *Syntax and Semantics of Spatial P*, 305–330, Amsterdam: John Benjamins. 223

PARTEE, B.H. & BORSCHEV, V. (2003). Genitives, relational nouns, and argument-modifier ambiguity. In E. Lang, C. Maienborn & C. Fabricius-Hansen, eds., *Modifying Adjuncts*. 162

PAYNE, J. (2004). Inflecting postpositions in Indic and Kashmiri. In F. Plank, ed., *Double Case: Agreement by Suffixaufnahme*, Oxford: Oxford University Press. 148

PLATTS, J. (1967). *A Grammar of the Hindustani or Urdu Language*. London: Crosby Lockwood and Son, 1909, republished in 2002 by Sang-e-Meel Publications, Lahore. 151, 188, 213

PLATZACK, C. (1982). Transitive adjectives in Swedish: a phenomenon with implications for the theory of abstract Case. *Linguistic Review*, **2**, 39–56. 177

POLLARD, C. & SAG, I.A. (1987). *Information-Based Syntax and Semantics*. CA: CSLI Publications. 61

POORNIMA, S. & KOENIG, J.P. (2008). Reverse complex predicates in Hindi. In *Proceedings of the 24th Northwest Linguistic Conference*. 28

PROCTER, P. (1987). *Longman dictionary of contemporary English*. Longman, London. 15

QUIRK, R. & MULHOLLAND, J. (1964). Complex prepositions and related sequences. *English studies*, **44**, 64–73. 238

RAPTSCHINSKY, B. (1946). *Russische Spraakkunst*. Groningen: Wolters, i edn. 3

RAZA, G. (2010a). Inferring Subcat frames of verbs in Urdu. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. 76

RAZA, G. (2010b). Semantic Realizations of the verb *ho* 'be/become' in Urdu. In *Proceedings of the First International Linguistic Conference in Warsaw (ILCW'10)*, Warsaw, Poland. 129

RAZA, G. (2010c). Analyzing the stucture of Urdu NPs with multiple genitives. In *Proceedings of the Conference on Language and Technology 2010 (CLT'10)*, Islamabad. 104, 148

RAZA, G. (2010d). *Reduction of Compound Adpositions in Persian, Urdu and Saraiki*. Presentation given at the Sixth International Contrastive Linguistics Conference (ICLC'06), Berlin. 253

RAZA, G. (2011). *Modeling Complex Adpositions in LFG: A case study form Persian and Urdu*. Presentation given at the Fourth International Conference on Iranian Linguistics (ICIL4), Uppsala. 252

RAZA, G. & AHMED, T. (2011a). Argument Scrambling within Urdu NPs. In M. Butt & T.H. King, eds., *Proceedings of the 16th LFG conference held in Hong Kong*, Stanford: CSLI Publications. 197

RAZA, G. & AHMED, T. (2011b). *Rendering Spatial Expressions involving Axial Parts in LFG*. Presentation given at the Thematic Workshop on Spatial and Temporal Relations in LFG, Debrecen. 226

RIZVI, S.M.J. (2007). *Development of Algorithms and Computational Grammar for Urdu*. Ph.D. thesis, Pakistan Institute of Engineering and Applied Sciences (PIEAS), Islamabad. 42, 53

RIZVI, S.M.J. (2008). Indications of Urdu tetravalent verbs having 'oblique agents' in the argument structure. In M. Butt & T.H. King, eds., *Proceedings of the LFG08 Conference*, CA: CSLI Publications. 19

ROEHRS, D. (2006). Split-NP as an Instance of Sideward Movement. University of Pennsylvania Working Papers in Linguistics 13.1, 1-14. 183

ROHEN WOLFF, S., MACLEOD, C. & MEYERS, A. (1994). COMPLEX word classes, manual. Linguistic Data Consortium, UPenn. 14

ROLAND, D. & JURAFSKY, D. (2002). Verb Sense and Verb Subcategorization Probabilities. In *The Lexical Basis of Sentence Processing: Formal, Computational, and Experimental Issues*, 325–346, John Benjamins, Amsterdam. 55

SADLER, L. & ARNOLD, D.J. (1994). Prenominal Adjectives and the Phrasal/Lexical Distinction. *Journal of Linguistics*, 187–226. 193

SAKSENA, A. (1982). *Topics in the Analysis of Causatives with an Account of Hindi Paradigms*. University of California Press, Berkeley. 19

SANFILIPPO, A. (1993). LKB encoding of lexical knowledge. In T. Briscoe, A. Copestake & V. de Paiva, eds., *Default inheritance in unification-based approaches to the lexicon*, Cambridge: Cambridge University Press. 14

SARKAR, A. & ZEMAN, D. (2000). Automatic Extraction of Subcategorization Frames for Czech. In *Proceedings of the 18th International Conference on Computational Linguistics*, 691–697, Saarbrücken, Germany. 88

SATYANARAYANA, P. (1982). Syntax of *hona* in Hindi. In O.N. Koul, ed., *Topics in Hindi Linguistics*, vol. 2, New Delhi. 127

Schmidt, R.L. (1999). *Urdu: An Essential Grammar*. London: Routledge. 188, 212

Schulte im Walde, S. (2006). Experiments on the Automatic Induction of German Semantic Verb Classes. *Computational Linguistics (32)2*, 159–194. 98

Schulte im Walde, S. & Brew, C. (2002). Inducing German Semantic Verb Classes from Purely Syntactic Subcategorisation Information. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 223–230, Philadelphia, PA. 98

Seiler, H. (1983). *Possession as an Operational Dimension of Language*. Tübingen: Günter Narr Verlag. 153

Sekerina, I. (1997). *The Syntax and Processing of Scrambling Constructions in Russian*. Ph.D. thesis, CUNY. 183

Simpson, J. (1983). *Aspects of Warlpiri Morphology and Syntax*. Ph.D. thesis, MIT. 182

Simpson, J. (1991). *Warlpiri Morpho-Syntax: A Lexicalist Approach*. Dordrect: Kluwer. 182

Subbarao, K.V. (1984). *Complementation in Hindi syntax*. Academic Series in Linguistics, I edn. 141

Sulger, S. (2011). A Parallel Analysis of have-type copular constructions in two have-less Indo-European Languages. In M. Butt & T.H. King, eds., *Proceedings of 16th LFG conference held in Hong Kong*, CSLI Publications. 112

Svenonius, P. (2004). Spatial P in English. Ms. University of Tromsø. http://ling.auf.net/lingBUZZ/000001. 207

Svenonius, P. (2006). The Emergence of Axial Parts. In P. Svenonius & M. Pantcheva, eds., *Nordlyd: Tromsø Working Papers on Language and Linguistics: 33.1, Special Issue on Adpositions*, 49–77, University of Tromsø, Tromsø. 201, 222, 227

Svenonius, P. (2007). Adpositions, particles, and the arguments they introduce. In E. Reuland, T. Bhattacharya & G. Spathas, eds., *Argument Structure*, 77–110, Amsterdam: John Benjamins. 201, 202, 221

Svenonius, P. (2008). Projections of P. In A. Asbury, J. Dotlacil, B. Gehrke & R. Nouwen, eds., *Syntax and Semantics of Spatial P*, 63–84, Amsterdam: John Benjamins. 201, 221, 223

ten Hacken, P., Maas, H. & Maegaard, B. (1991). Dictionaries in eurota. In C. Copeland, J. Durand, S. Krauwer & B. Maegaard, eds., *The Eurota linguistic specifications*, The Commission of the European Communities, Brussels. 15

Travis, L. (1988). The Syntax of Adverbs. In *McGill Working Papers in Linguistics: Proceedings of the IVth Workshop on Comparative Germanic Syntax*, 280–310. 193

Trawinski, B. (2003). Licensing Complex Prepositions via Lexical Constraints. In F. Bond, A. Korhonen, D. McCarthy & A. Villavicencio, eds., *Proceedings of the AICL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, 97–104, Sapporo Convention Center, Sapporo, Japan. 238, 252

Ushioda, A., Evans, D.A., Gibson, T. & Waibel, A. (1993). The automatic acquisition of frequencies of verb subcategorization frames from tagged corpora. In *Proceedings of the Workshop on the Acquisition of Lexical Knowledge from Text*, 95–106, Columbus, OH. 14, 56, 58, 61

van Hoof, H. (1997). Left Dislocation and Split Topics. In E. Anagnostopoulou, H. van Riemsdijk & F. Zwarts, eds., *Materials on Left Dislocation*, vol. 14 of Liguistic Aktuell, 275–305, Amsterdam and Philadelphia: John Benjamins. 183

van Riemsdijk, H. (1978). A case study in syntactic markedness: the binding nature of prepositional phrases. Lisse: Peter de Ridder. 207

van Riemsdijk, H. (1989). Movement and Regeneration. In *Dialectal Variation and Theory of Grammar*, 105–136, Dordrecht: Foris. 183

van Riemsdijk, H. (1990). Functional prepositions. In H. Pinkster & I. Genee, eds., *Unity in diversity: papers presented to Simon Dik on his 50th birthday*, 229–241, Foris: Dordrecht. 207

Vanocchi, M., Rosini, R., Carenini, M., Prodanof, I. & Calzolari, N. (1994). Developing a neutral formalism for verbal representation. Technical Report ILC-NLP-1994-I, ILC-CNR, Pisa. 15

Vendler, Z. (1968). *Linguistics in Philosophy*. Ithaca: Cornell University Press. 127

Williams, E. (1982). Another Argument that Passive is Transformational. *Linguistic Inquiry*, **13**, 160–163. 191

Windfuhr, L.G. (1979). *Persian Grammar: History and State of its Study*. Mouton Publishers, The Hague. 31

WINTER, Y. (2001). Measure phrase modification in a vector space semantics. In K. Megerdoomina & L.A. Bar-el, eds., *Proceedings of WCCFL 20*, 607–620, Somerville, MA: Cascadilla Press. 219

ZWARTS, J. (1995). The semantics of relative position. *Semantics and Linguistic Theory*, **5**, 405–422. 218

ZWARTS, J. (2000). Vectors across spatial domains: From place to size, orientation, shape and parts. In E. van der Zee & J. Slack, eds., *Axes and Vectors in Language and Space*, Oxford: Oxford University Press. 218, 220

ZWARTS, J. & WINTER, Y. (1997). A semantic characterization of locative PPs. *Semantics and Linguistic Theory*, **7**, 294–311. 218, 219

ZWICKY, A. (1987). Suppressing the Z's. *Journal of Linguistics*, **23**. 149